


```

; APPLICANT: STABLETON, JACK T.
; TITLE OF INVENTION: A PROTEIN THAT INTERACTS WITH LIPIDS AND
; FILE REFERENCE: IOWA:045W0
; CURRENT APPLICATION NUMBER: PCT/US03/19834
; PRIOR FILING DATE: 2003-05-27
; PRIOR APPLICATION NUMBER: 10/445,724
; PRIOR FILING DATE: 2002-06-28
; PRIOR FILING DATE: 2002-06-28
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; TYPE: PRT
; ORGANISM: Hepatitis C virus
PCT-US03-19834-2

```

```

Query Match          14.0%; Score 254; DB 1; Length 3011;
Best Local Similarity 31.8%; Pred. No. 2; Le-14;
Matches 85; Conservative 27; Mismatches 77; Indels 78; Gaps 10;

```

```

QY 23 RPAVTPREVLYREFDEMEBAS-HLPYIEGQMLAEQFKOKALAF---ASRQNHVSPTHY 78
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1691 KPAITPDEVLYREFDEMECSQHPYIEGQMLAEQFKOKALGLQTSRQAEV----- 1745
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 79 VPESDAPPAITPDEVLYREFDEMEBASR-GNHVSPTHYVPESDASQAAPYIEO---- 133
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1746 ----IAPAV-----QTMQKLETFWAKHMMNFISSGIQLVGLSTLPGNPAIASLMAF 1793
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 134 -AOVIAHQEKEKVLAF-----ASRGNHDSPTHYVPESDAKPAI----- 170
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1794 TAAVTSPLTTSQTLFNILGQVVAQLAAGAAATAVGAGLGAIGSVGLKVLIDILA 1853
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 171 -----IPDEVLYREFDEMEESQHL--PYIEGQML-----AE 201
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1854 GYAGVAGALVAKKINGSEVPSTEDLVNLPALISGALVGVVCAILRRHVGPGECAV 1913
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 202 QFKOKALAFASRGNHVAPTHYVTSDA 228
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1914 QMNRILAFASRGNHVSPTHYVTSDA 1940
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 3
US-10-370-481-31
; Sequence 31, Application US/10370481
; GENERAL INFORMATION:
; APPLICANT: Hitachi, Ltd.
; TITLE OF INVENTION: AIP-INTERACTING PROTEINS AND USE THEREOF
; FILE REFERENCE: HITA.0164
; CURRENT APPLICATION NUMBER: US/10/370,481
; PRIOR FILING DATE: 2003-02-24
; PRIOR APPLICATION NUMBER: US 60/358,369
; PRIOR FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 60/019,808
; PRIOR FILING DATE: 2002-09-03
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 31
; LENGTH: 4574
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-370-481-31

```

```

Query Match          6.2%; Score 113.5; DB 6; Length 4574;
Best Local Similarity 20.1%; Pred. No. 0.13;
Matches 73; Conservative 65; Mismatches 169; Indels 57; Gaps 10;

```

```

QY 14 AEOFKOKAL-----RPAVTPREVLYREFDEMEASHLPYIEGQML 55
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 2013 AEEVQKSLAAEEBAAROKKALEVERLKVANVEEARLREBAEESARQLQAEAAOK 2072
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

QY 56 AEOFKOKALAFASRGNHVSPTHYVPESDAPPAITPDEVLYREFDEME-EAFASR--GNH 112
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 2073 RLQAEKKAHAFVQK-----EQLQOOLQOESVL-----DQLGEEBAARRAAE 2119
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 113 VSPAHYVPESDASQAAPYIEQAOVIAHQFKEKVLAF--SRGNHDSPTHYVPESDAKPAI 171
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 2120 AEEARVQAEAREAAQRQVEEAEERLKQAEQQAQAQAQAAEKLRBAEQAARRQA 2179
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 172 PDREVLYREFDEMECSQHPYIEGQMLAEQFKOKALAFASRGNHVAPTHYVTSDAKPA 231
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 2180 EQAALROKQADAEEMEKHKFAEQTLRQKAVEDELTLRLQ-----LEETDHQKN 2230
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 232 LVPDKEVLYQOYDEMEBAFASRGNHVAPTHYVPESDASASLPYMDETRAIAGQFKEKVL 291
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 2231 LL-----DEELQRLKAEATBAARQSOVEEELFSVRYQMELSKLRARIEAENRA 2280
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 232 FASRGNHVSPTHYVPESQVAVTPDKEILYEAPDE---MEBASKALIEGQMAEMLK 348
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 2281 LILRKQNTQRFLOEAEKMKQVAAEAAARLSVAQEAARLRQLAEEDLAQORALAEKMLK 2340
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 349 SKIQ 352
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 2341 EXMQ 2344
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 4
US-60-479-073-495
; Sequence 495, Application US/60479073
; GENERAL INFORMATION:
; APPLICANT: De Wilde, Gert Jules Hector
; APPLICANT: Saunders, Michael John Scott
; APPLICANT: Logghe, Marc Georges
; TITLE OF INVENTION: Amino acid sequences useful for developing compounds for the
; FILE REFERENCE: D00590.70042.US
; CURRENT APPLICATION NUMBER: US/60/479,073
; PRIOR FILING DATE: 2003-06-17
; NUMBER OF SEQ ID NOS: 526
; SOFTWARE: Patent In version 3.2
; SEQ ID NO 495
; LENGTH: 4574
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-479-073-495

```

```

Query Match          6.2%; Score 113.5; DB 7; Length 4574;
Best Local Similarity 20.1%; Pred. No. 0.13;
Matches 73; Conservative 65; Mismatches 169; Indels 57; Gaps 10;

```

```

QY 14 AEOFKOKAL-----RPAVTPREVLYREFDEMEASHLPYIEGQML 55
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 2013 AEEVQKSLAAEEBAAROKKALEVERLKVANVEEARLREBAEESARQLQAEAAOK 2072
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 56 AEOFKOKALAFASRGNHVSPTHYVPESDAPPAITPDEVLYREFDEME-EAFASR--GNH 112
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 2073 RLQAEKKAHAFVQK-----EQLQOOLQOESVL-----DQLGEEBAARRAAE 2119
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 113 VSPAHYVPESDASQAAPYIEQAOVIAHQFKEKVLAF--SRGNHDSPTHYVPESDAKPAI 171
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 2120 AEEARVQAEAREAAQRQVEEAEERLKQAEQQAQAQAQAQAAEKLRBAEQAARRQA 2179
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 172 PDREVLYREFDEMECSQHPYIEGQMLAEQFKOKALAFASRGNHVAPTHYVTSDAKPA 231
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 2180 EQAALROKQADAEEMEKHKFAEQTLRQKAVEDELTLRLQ-----LEETDHQKN 2230
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 232 LVPDKEVLYQOYDEMEBAFASRGNHVAPTHYVPESDASASLPYMDETRAIAGQFKEKVL 291
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 2231 LL-----DEELQRLKAEATBAARQSOVEEELFSVRYQMELSKLRARIEAENRA 2280
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 232 FASRGNHVSPTHYVPESQVAVTPDKEILYEAPDE---MEBASKALIEGQMAEMLK 348
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 2281 LILRKQNTQRFLOEAEKMKQVAAEAAARLSVAQEAARLRQLAEEDLAQORALAEKMLK 2340
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

QY 349 SKIO 352
DB 2341 EKXQ 2344

RESULT 5
US-60-479-073-497

Sequence 497, Application US/60479073
GENERAL INFORMATION:
APPLICANT: De Wilde, Gert Jules Hector
APPLICANT: Saunders, Michael John Scott
TITLE OF INVENTION: Amino acid sequences useful for developing compounds for the
TITLE OF INVENTION: prevention and/or treatment of metabolic diseases and nucleotide
FILE REFERENCE: D00590.70042.US
CURRENT APPLICATION NUMBER: US/60/479, 073
CURRENT FILING DATE: 2003-06-17
NUMBER OF SEQ ID NOS: 526
SOFTWARE: Patentin version 3.2
SEQ ID NO 497
LENGTH: 4684
TYPE: PRT
ORGANISM: Homo sapiens
US-60-479-073-497

Query Match 6.2%; Score 113.5; DB 7; Length 4684;

Best Local Similarity 19.7%; Pred. No. 0.13;
Matches 71; Conservative 63; Mismatches 176; Indels 51; Gaps 8;

QY 14 AEOFKOKAL-----RPVLPDREV-----YQFDEMEASHLPYIEQGMQL 55
DB 2123 AEEVVKSLAEAEARORAKALEVEERLAKVEEARSLERAEQESARQLQAQEAOK 2182
QY 56 AEFKOKALAFASRGNHVSTHYVPSDARPAIIPREVLRHREFDEMEAFASRGNHVSP 115
DB 2183 RLQAEKHAFAVQOK-----EOELQOTLQOOSVLDRLRGEAEAA-RRAAEEAE 2232
QY 116 AHVPESDASQAPYIEQAQVIAHQFEKVLFA-SRGNHDSPTHYVPSDARPAIIPDR 174
DB 2233 ARVQAREAAQSRQVEERLQKSAEQAQARAQAAAEKLRKAEQAARAAQEAQ 2292
QY 175 EVLYREFDEMESSOHLPYIEQGMMLAEQFKOKALAFASRGNHVAPTHYVTESDAKALVP 234
DB 2293 ALRQKQADAEEMKHKFAEQTLRQQAQVEQLTLRLQ-----LEETDQKNLL- 2342
QY 235 DKEVLYQOYDEMEAFASRGNHVAPTHYVTESDASASLPRMDETRAIAGQFEKVLAFAS 294
DB 2343 -----DEELQRLKAEATEAARORSQVEEELFSVRVQMEELSKLKARIEAENRALIL 2393
QY 295 RGNHVSPTHYVPESEPOVVVTPDKETLYEAFDE---MEASKAALIEEQRMAMEMKSKI 351
DB 2394 RKDNDTORFLQEAEMKQVAAEAAARLSVAQEAARLRQLAEEDLQAQALAEKMKEXM 2453
QY 352 Q 352
DB 2454 Q 2454

RESULT 6
US-60-479-073-506
Sequence 506, Application US/60479073
GENERAL INFORMATION:
APPLICANT: De Wilde, Gert Jules Hector
APPLICANT: Saunders, Michael John Scott
TITLE OF INVENTION: Amino acid sequences useful for developing compounds for the
TITLE OF INVENTION: prevention and/or treatment of metabolic diseases and nucleotide
sequences encoding such amino acid sequences.
US-60-479-073-506

NUMBER OF SEQ ID NOS: 526
SOFTWARE: Patentin version 3.2
SEQ ID NO 506
LENGTH: 4684
TYPE: PRT
ORGANISM: Homo sapiens
US-60-479-073-506

Query Match 6.2%; Score 113.5; DB 7; Length 4684;
Best Local Similarity 19.7%; Pred. No. 0.13;
Matches 71; Conservative 63; Mismatches 176; Indels 51; Gaps 8;

QY 14 AEOFKOKAL-----RPVLPDREV-----YQFDEMEASHLPYIEQGMQL 55
DB 2123 AEEVVKSLAEAEARORAKALEVEERLAKVEEARSLERAEQESARQLQAQEAOK 2182
QY 56 AEFKOKALAFASRGNHVSTHYVPSDARPAIIPREVLRHREFDEMEAFASRGNHVSP 115
DB 2183 RLQAEKHAFAVQOK-----EOELQOTLQOOSVLDRLRGEAEAA-RRAAEEAE 2232
QY 116 AHVPESDASQAPYIEQAQVIAHQFEKVLFA-SRGNHDSPTHYVPSDARPAIIPDR 174
DB 2233 ARVQAREAAQSRQVEERLQKSAEQAQARAQAAAEKLRKAEQAARAAQEAQ 2292
QY 175 EVLYREFDEMESSOHLPYIEQGMMLAEQFKOKALAFASRGNHVAPTHYVTESDAKALVP 234
DB 2293 ALRQKQADAEEMKHKFAEQTLRQQAQVEQLTLRLQ-----LEETDQKNLL- 2342
QY 235 DKEVLYQOYDEMEAFASRGNHVAPTHYVTESDASASLPRMDETRAIAGQFEKVLAFAS 294
DB 2343 -----DEELQRLKAEATEAARORSQVEEELFSVRVQMEELSKLKARIEAENRALIL 2393
QY 295 RGNHVSPTHYVPESEPOVVVTPDKETLYEAFDE---MEASKAALIEEQRMAMEMKSKI 351
DB 2394 RKDNDTORFLQEAEMKQVAAEAAARLSVAQEAARLRQLAEEDLQAQALAEKMKEXM 2453
QY 352 Q 352
DB 2454 Q 2454

RESULT 7
US-10-273-573-10712
Sequence 10712, Application US/10273573
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc
TITLE OF INVENTION: NOVEL MACROPHAGE NUCLEIC ACIDS AND POLYPEPTIDES
FILE REFERENCE: 21272-066
CURRENT APPLICATION NUMBER: US/10/273, 573
CURRENT FILING DATE: 2002-10-18
PRIOR APPLICATION NUMBER: 09/522, 929
PRIOR FILING DATE: 2000-04-18
PRIOR APPLICATION NUMBER: 09/770, 160
PRIOR FILING DATE: 2001-01-26
NUMBER OF SEQ ID NOS: 10994
SOFTWARE: Custom
SEQ ID NO 10712
LENGTH: 365
TYPE: PRT
ORGANISM: Homo sapiens
US-10-273-573-10712

Query Match 5.8%; Score 106; DB 6; Length 365;

Best Local Similarity 20.4%; Pred. No. 0.016;
Matches 64; Conservative 41; Mismatches 100; Indels 108; Gaps 13;

QY 69 RGNHVSPTHYVPSDAR-----PAIIPDREVLRHREFDEMEAFASRGNHVSPAHYVPS 122
DB 2 REXSDPNIVLESIRKHSPESSIVESV-----RKHSPESSIVLES 47
QY 123 DASQAPYIEQAQVIAHQFE---KVLAFASRGNHDSPTHYVPSDAKPAIIPREVLYR 179
DB 48 -----YVSHGSELSHRSVLESYVRDKHSPESSIVLES-----YVRDK----- 84

Mon Jul 28 11:16:34 2003

OY 180 EFDEMEESQHLPIYIEGGMMLAEQFKOKALAFASRGNHVAPTHVYES---DAKPAIVPDK 236
DB 85 -----HSEPPSSVLEBSF-----VRDKHSEPPSSVLESTECSSPPAM--EK 122
OY 237 EVLYOQYDEMEEAFA-----ASRGNHVAPTHVVE-----SDASASLPYMDT 278
DB 123 SWAENDFPELREGFRNSVYSELQIEQKKEVKNFKENKLOECISRIINTEKCKELMEL 182
OY 279 RAIGQFKKYLAFASRGNHVSPHYVESEPOVVVTPDKELIYEAPEDEMEASKAL-I 337
DB 183 KAKRELEECRRFRFRSCDQLEERDIIQENSFNI-----ARQANIQI 224
OY 338 EGGORMAMMLKSK 350
DB 225 QEIORMFORYSLK 237

RESULT 8
US-60-479-073-498
; Sequence 498, Application US/60479073
; GENERAL INFORMATION:
; APPLICANT: De Wilde, Gert Jules Hector
; APPLICANT: Saunders, Michael John Scott
; APPLICANT: Logghe, Marc Georges
; TITLE OF INVENTION: Amino acid sequences useful for developing compounds for the
; TITLE OF INVENTION: prevention and/or treatment of metabolic diseases and nucleotide
; TITLE OF INVENTION: sequences encoding such amino acid sequences.
; FILE REFERENCE: D00590.70042.US
; CURRENT APPLICATION NUMBER: US/60/479.073
; CURRENT FILING DATE: 2003-06-17
; NUMBER OF SEQ ID NOS: 526
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 498
; LENGTH: 5373
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-479-073-498

Query Match 5.7%; Score 103; DB 7; Length 5373;
Best Local Similarity 18.5%; Pred. No. 1.4;
Matches 78; Conservative 73; Mismatches 132; Indels 138; Gaps 19;

OY 4 IPYLEGGMMLAEQFKOKALRPVIPPREVLYOEFDMEEBASH-----LPY----- 48
DB 3584 IFOFOQ-----ROKELKKEVMEHRLVL-----DTVNEVSRLLELVPRAREGLDKLV 3631
OY 49 -----IEGGMMLAEQFKOKA---LAFASRGNH-----VSPTHVPE 81
DB 3632 SDANEQYKLVSDTIGRVDEIDALIQSQYEQADALAWVATETKRLMALGPTRL--E 3689
OY 82 SDARPAIIPREV---LHREFDEMEEAFAFASRGN-----HVSAPHVPESDA 124
DB 3690 QDQTTAOLQVQKAFSIDIRHKDSMDLFSHRSIEIFGTGCEBQKTVLOEKTESLIQYEA 3749
OY 125 -----SQAAPYIEQAQVIAHQFKEKYLAFASRGNHSDPTHVPESSAKPAIIP---DRE 175
DB 3750 ISLNSERVARLERAAQVLVNOFWE-----TYEELSPWIEETRALIAQLPSPAIDHE 3800
OY 176 VLYPEFDEMEE---SQHLPIYIEGGMMLAEQFKOKALAFASRGNHVAPTHVYESDAKP 230
DB 3801 QLRQOQDEMRQLRESIAEHKPHIDKLKIGPOLKE----- 3835
OY 231 ALVPDK-EVLYOQYDEMEEAFAFASRGNHVAPTHVYESDASASLPYMDETRAIAGQFKEXV 289
DB 3836 -LNPDEGEVVEEKYQKAEENMYAQIKEEVQRALALDEAVSQSTQIT-----EFHDKI 3886
OY 290 LAFASRGNHVSPHYVESEPOVVVTPDKELIYEAPEDEMEASKALIEGGMAMMLKSK 349
DB 3887 EPMLETLENLSSRLRMP---PLIPAEVDK--IRECISDNKSAI--VELEKLOPSFEALKR 3939
OY 350 K 350

DB 3940 R 3940

RESULT 9
US-60-479-073-500
; Sequence 500, Application US/60479073
; GENERAL INFORMATION:
; APPLICANT: De Wilde, Gert Jules Hector
; APPLICANT: Saunders, Michael John Scott
; APPLICANT: Logghe, Marc Georges
; TITLE OF INVENTION: Amino acid sequences useful for developing compounds for the
; TITLE OF INVENTION: prevention and/or treatment of metabolic diseases and nucle
; TITLE OF INVENTION: sequences encoding such amino acid sequences.
; FILE REFERENCE: D00590.70042.US
; CURRENT APPLICATION NUMBER: US/60/479.073
; CURRENT FILING DATE: 2003-06-17
; NUMBER OF SEQ ID NOS: 526
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 500
; LENGTH: 5430
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-479-073-500

Query Match 5.7%; Score 103; DB 7; Length 5430;
Best Local Similarity 18.5%; Pred. No. 1.4;
Matches 78; Conservative 73; Mismatches 132; Indels 138; Gaps 19;

OY 4 IPYLEGGMMLAEQFKOKALRPVIPPREVLYOEFDMEEBASH-----LPY----- 48
DB 3640 IFOFOQ-----ROKELKKEVMEHRLVL-----DTVNEVSRLLELVPRAREGLDKLV 3687
OY 49 -----IEGGMMLAEQFKOKA---LAFASRGNH-----VSPTHVPE 81
DB 3688 SDANEQYKLVSDTIGRVDEIDALIQSQYEQADALAWVATETKRLMALGPTRL--E 3745
OY 82 SDARPAIIPREV---LHREFDEMEEAFAFASRGN-----HVSAPHVPESDA 124
DB 3746 QDQTTAOLQVQKAFSIDIRHKDSMDLFSHRSIEIFGTGCEBQKTVLOEKTESLIQYEA 3805
OY 125 -----SQAAPYIEQAQVIAHQFKEKYLAFASRGNHSDPTHVPESSAKPAIIP---DRE 175
DB 3806 ISLNSERVARLERAAQVLVNOFWE-----TYEELSPWIEETRALIAQLPSPAIDHE 3856
OY 176 VLYPEFDEMEE---SQHLPIYIEGGMMLAEQFKOKALAFASRGNHVAPTHVYESDAKP 230
DB 3857 QLRQOQDEMRQLRESIAEHKPHIDKLKIGPOLKE----- 3891
OY 231 ALVPDK-EVLYOQYDEMEEAFAFASRGNHVAPTHVYESDASASLPYMDETRAIAGQFKEXV 289
DB 3892 -LNPDEGEVVEEKYQKAEENMYAQIKEEVQRALALDEAVSQSTQIT-----EFHDKI 3942
OY 290 LAFASRGNHVSPHYVESEPOVVVTPDKELIYEAPEDEMEASKALIEGGMAMMLKSK 349
DB 3943 EPMLETLENLSSRLRMP---PLIPAEVDK--IRECISDNKSAI--VELEKLOPSFEALKR 3995
OY 350 K 350
DB 3996 R 3996

RESULT 10
US-10-333-120A-10
; Sequence 10, Application US/10333120A
; GENERAL INFORMATION:
; APPLICANT: Januliczky, Robert
; APPLICANT: Tammel, Francesco
; APPLICANT: Sjolholm, Anders
; APPLICANT: Pozzi, Gianni
; APPLICANT: Bjorck, Lars
; TITLE OF INVENTION: PROTEIN
; FILE REFERENCE: 13519PCTUS
; CURRENT APPLICATION NUMBER: US/10/333,120A

CURRENT FILING DATE: 2002-01-14
PRIOR APPLICATION NUMBER: PCT/EP01/08409
PRIOR FILING DATE: 2001-07-19
PRIOR APPLICATION NUMBER: SE 0002728-4
PRIOR FILING DATE: 2000-07-20
NUMBER OF SEQ ID NOS: 11
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10
LENGTH: 693
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-10-333-120A-10

Query Match 5.2%; Score 95; DB 6; Length 693;
Best Local Similarity 22.0%; Pred. No. 0.38;
Matches 55; Conservative 49; Mismatches 122; Indels 24; Gaps 7;

QY 15 EGFKOKALRAVIVPDRREVLYOEFDME--EASHLYIEGQMLAEQFKALAFASRQ 70
DB 230 ELYKVKANER--DEQKIQAEAEVESKQAEATRLKKTIDREAE--EKKRADAKEQ 285
QY 71 NHVSPHYVPESDAPAIIPDREVLYHREFDEMEAFASRGNHVSFAHYV--PESDASQAP 129
DB 286 GK-----PKGRAKRGVPELATPDCKENDAKSSDSSVGEETLPSPLKPEKKVLAEK 338
QY 130 YIQAOVIYAHQFEKVLAFASRGNHDSPTHYVPESDAPAIIPDREVLYHREFDEMEESQH 189
DB 339 KVEAKKAKEDQEE-----DRNYPNTYKTLLELAES--DVEYKAELELVKEAK 390
QY 190 LPYIEGQMLAEQFKOKALAFASRGNHVAPTHYVTESDAPALVPDEVLVYQYDMEEA 249
DB 391 EPNNEKVQAKAEVESKKAETRLKKTIDRKKAEEBAKKAEDVKKEKAEQOPA 450
QY 250 FASRGNHVP 259
DB 451 PAPKAEKAP 460

RESULT 11
US-10-391-363A-61
Sequence 61, Application US/10391363A
GENERAL INFORMATION:
APPLICANT: Monsanto Technology LLC
APPLICANT: Valentin, Henry E
APPLICANT: Venkatesh, Tyamagondlu V
APPLICANT: Karunananda, Balasulojini
TITLE OF INVENTION: Homogentisate Prenyl Transferase ("HPT") Nucleic Acids and Polypeptide
TITLE OF INVENTION: and Uses Thereof
FILE REFERENCE: REN-02-052 MON-52933
CURRENT FILING DATE: 2003-03-18
PRIOR APPLICATION NUMBER: US/10391,363A
PRIOR FILING DATE: 2002-03-19
NUMBER OF SEQ ID NOS: 95
SOFTWARE: PatentIn version 3.1
SEQ ID NO 61
LENGTH: 970
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-10-391-363A-61

Query Match 5.2%; Score 95; DB 6; Length 970;
Best Local Similarity 21.4%; Pred. No. 0.62;
Matches 72; Conservative 47; Mismatches 123; Indels 94; Gaps 14;

QY 31 EVLYOEFDMEASHLYIEGQMLAEQFKALAFASRGNHVSPTHYVP 80
DB 329 ELYKVKANER--DEQKIQAEAEVESKQAEATRLKKTIDREAE--EKKRADAKEQ 285
QY 71 NHVSPHYVPESDAPAIIPDREVLYHREFDEMEAFASRGNHVSFAHYV--PESDASQAP 129
DB 286 GK-----PKGRAKRGVPELATPDCKENDAKSSDSSVGEETLPSPLKPEKKVLAEK 338
QY 130 YIQAOVIYAHQFEKVLAFASRGNHDSPTHYVPESDAPAIIPDREVLYHREFDEMEESQH 189
DB 339 KVEAKKAKEDQEE-----DRNYPNTYKTLLELAES--DVEYKAELELVKEAK 390
QY 190 LPYIEGQMLAEQFKOKALAFASRGNHVAPTHYVTESDAPALVPDEVLVYQYDMEEA 249
DB 391 EPNNEKVQAKAEVESKKAETRLKKTIDRKKAEEBAKKAEDVKKEKAEQOPA 450
QY 250 FASRGNHVP 259
DB 451 PAPKAEKAP 460

QY 122 S-----DAS--QAAPYIEQOVAHQFE-----KYLAFASRGNHDSPTHYVP 162
DB 384 DKLSKEVPAAMVKAAYKASEAIDNVNEBGLAKQKLEIFGEFTKEGDNNSADDLSE 443
QY 163 ESDAKPAIIPDREVLYHREFDEMEESQHPY-----IEGQMLAEQFKALAFASRGN 215
DB 444 RTTKAPVYVMDAEEBDSVETKOTRTYETTRGLEIENEMVIDETKST----- 495
QY 216 HVAPTHYVTESDAPALVPDKEVLVYQYDMEEAFAFRGNHVAPTHYVTESD-- 267
DB 496 -----ETKTEAPSRIVMDKE-----GDEITKSTETETEPASRIVMETEODETMS 542
QY 268 -ASASLPYMDETRAIQFKEKVLAFASRGNHVSPT 302
DB 543 RARS-----DEAALIS--KSSQVASMELISQSPR 571

RESULT 12
PCT-US02-18638A-2
Sequence 2, Application PCT/TUS0218638A
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc. et al.
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
TITLE OF INVENTION: OF CERVICAL CANCER
FILE REFERENCE: MRI-035PC
CURRENT FILING DATE: 2002-06-12
PRIOR APPLICATION NUMBER: PCT/US02/18638A
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/298,159
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/298,155
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/335,936
PRIOR FILING DATE: 2001-11-14
NUMBER OF SEQ ID NOS: 238
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 3907
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US02-18638A-2

Query Match 5.2%; Score 95; DB 1; Length 3907;
Best Local Similarity 19.3%; Pred. No. 4.6;
Matches 68; Conservative 54; Mismatches 145; Indels 86; Gaps 14;

QY 19 QALRAVIVPDRREVLYOEFDMEEA--SHLYIEGQMLAEQFKALAFASRGNHVSPT 76
DB 2020 QKQVLAEL--DVEQVSRFIELEOEKNTLMROONQALEKMRKFLDEQ--AIDRE 2077
QY 77 HVPESDAPAIIPDREVLYHREFDEMEEAFAFRGNHVSFAHYV--PESDASQAPYIEQOV 136
DB 2078 H-----ENDVFOQETQKLEQOL-----KVP--RQPISEHO-----TREVOQ 2113
QY 137 IAHQFEKVLAFASRGNHDSPTHYVPESDAPAIIPDREVLYHREFDEMEESQHPYIEGQ 196
DB 2114 LANHLKEX-----IDKCELLLSKEQLORDQERNEE-----IEVL 2149
QY 197 MMLAEQFKOKALAFASRGNHVAPTHYVTESDAKPALV-----PDKEV--LY 240
DB 2150 EFRVLELEQALLVSADTEQKVEDRKHFGAVEAKPELSLEVQLAERDAIDRKEKEITNLE 2209
QY 241 QOYDMEEAFAFRGNHVAPTHYVTESDASASLPYMDETRAIQFKEKVLAFASRGNH 298
DB 2210 EQLQFRELLENKNEEVQOLHMQLEITQKKESTTRLOEIOENKLRKDMKEKGLAIK-- 2266
QY 299 VSPRHYVTESDAPALVPDKEVLVYQYDMEEAFAFRGNHVAPTHYVTESDAKPALV 348
DB 2267 -----ESDANSTODQHLVFGKFAQIIOEKVEIDQNEQVTKLOQOLK 2309

RESULT 13
PCT-US02-18638A-6

Query Match	5.2%	Score 95;	DB 1;	Length 3925;
Best Local Similarity	19.3%;	Pred. No. 4.6;		
Matches 68;	Conservative 54;	Mismatches 145;	Indels 86;	Gaps 14;

RESULT 14
US-10-273-573-9038

```

; SEQ ID NO 9038
; LENGTH: 2117

```

```
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
```

Query Match	5.2%	Score 94.5;	DB 6;	Length 217;
Best Local Similarity	20.5%;	Pred. No. 2.1;		
Matches 74; Conservative	61;	Mismatches 141;	Indels 85;	Gaps 17

RESULT 15
US-10-273-573-9034
; Sequence 9034, Application US/10273573

Query Match	5.2%	Score 94.5;	DB 6;	Length 2121;
Best Local Similarity	20.5%			
Mismatches 74;	Conservative 61;	Mismatches 141;	Gaps 17	

OY
QY

13 LAEQFKOKALBPVIPPREVLYQEFDMEASHPYEOM-----QLAEQFKOKALAF-66

DG
DB

530 LKQAQAEKQAQ-----LAOTLQQEQAAS-----QGRLHQVEQLSSSLKKKEEDOLK 574

```

QY 67 -----ASRONHVSPTHYVESDARPAIIPDREVLHREFDEMEEAFASRGNHVSPAHYV 119
Db 575 EVAEKQEXTRODHAQ--OLATAEERREASLREDDALKOLELEKAKLEILQOOLQV 632
QY 120 PESDASQAPYIEQAOVIAHQFKEV-----LAFASRGNHDSPTHYV-----ESDAKP 168
Db 633 ANEARDSDQTSVTQHQREKAELSRKEVEELQACVETARQOEHOAOVAELELOLRSEQOK 692
QY 169 AIPDREVLHREFDEMEESCHLPIYIEQGM-----MLAQFKOKALAFASRGNHVAPTHYV 223
Db 693 A--TEKERVNAQEKDQLOE--QLQALKESELKVTXGSLBEKRRADDALEEQRCISELKAE 748
QY 224 TESDAKPALVPDKVELYQOYDEMEEAFASRGNHVAPTHYVESDASASLPYMDETRAIAG 283
Db 749 TRS-----LVEOHK--RERKELEERAGR-----KGLEARLQOLGGEAHOAET 788
QY 284 QFKEKVLAFASRGNHVSPRHVPESEPOVVVTPDKEL-----YEAFDMEEASKALLIE 338
Db 789 EVLRRELAEMAAQOHTA-----ESECEQLV--KEVAAMRERYED--SQOEAOYGAMFQ 838
QY 339 E 339
Db 839 E 839

```

Search completed: July 25, 2003, 13:20:07
 Job time : 15 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 13:09:51 ; Search time 376 Seconds

(without alignments)
814,803 Million cell updates/sec

Title: US-09-491-146A-52

Sequence: 1 AAHPYLEGGMHAEQFKK.....KALIEEGQMAEMKSKITQ 352

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 5580241 seqs, 870357830 residues

Total number of hits satisfying chosen parameters: 5580241

Minimum DB seq length: 0
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending Patents: AA Main:*

1: /cgn2_6/ptodata/1/paa/PCUTS_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US097_COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US097b_COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US098b_COMB.pep.*
25: /cgn2_6/ptodata/1/paa/US099_COMB.pep.*
26: /cgn2_6/ptodata/1/paa/US099b_COMB.pep.*
27: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
28: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
29: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
30: /cgn2_6/ptodata/1/paa/US103_COMB.pep.*
31: /cgn2_6/ptodata/1/paa/US104_COMB.pep.*
32: /cgn2_6/ptodata/1/paa/US105_COMB.pep.*
33: /cgn2_6/ptodata/1/paa/US106_COMB.pep.*
34: /cgn2_6/ptodata/1/paa/US107_COMB.pep.*
35: /cgn2_6/ptodata/1/paa/US108_COMB.pep.*
36: /cgn2_6/ptodata/1/paa/US109_COMB.pep.*
37: /cgn2_6/ptodata/1/paa/US110_COMB.pep.*
38: /cgn2_6/ptodata/1/paa/US111_COMB.pep.*
39: /cgn2_6/ptodata/1/paa/US112_COMB.pep.*
40: /cgn2_6/ptodata/1/paa/US113_COMB.pep.*
41: /cgn2_6/ptodata/1/paa/US114_COMB.pep.*
42: /cgn2_6/ptodata/1/paa/US115_COMB.pep.*
43: /cgn2_6/ptodata/1/paa/US116_COMB.pep.*
44: /cgn2_6/ptodata/1/paa/US117_COMB.pep.*
45: /cgn2_6/ptodata/1/paa/US118_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Query Length DB ID Description
18 US-09-491-146-52 Sequence 52, Appl
US-09-491-146A-52 Sequence 52, Appl

3 512.5 28.2 829 23 US-09-881-239-5 Sequence 5, Appl
4 512.5 28.2 839 22 US-09-791-537-91118 Sequence 91118, A
5 512.5 28.2 1099 23 US-09-881-654-4 Sequence 4, Appl
6 257 14.1 3011 8 US-08-417-478-1 Sequence 1, Appl
7 257 14.1 3011 8 US-08-453-613-1 Sequence 1, Appl
8 256 14.1 492 22 US-09-791-537-79037 Sequence 79037, A
9 256 14.1 2013 3 US-07-635-451-10 Sequence 10, Appl
10 256 14.1 2013 3 US-07-769-996B-12 Sequence 12, Appl
11 256 14.1 2013 3 US-07-769-996C-12 Sequence 12, Appl
12 256 14.1 2013 4 US-08-099-706-12 Sequence 12, Appl
13 256 14.1 2013 7 US-08-383-682-10 Sequence 10, Appl
14 256 14.1 2201 26 US-10-085-476-2 Sequence 2, Appl
15 256 14.1 2620 3 US-07-769-996B-12 Sequence 32, Appl
16 256 14.1 2620 3 US-07-769-996C-12 Sequence 32, Appl
17 256 14.1 2620 4 US-08-099-706-32 Sequence 32, Appl
18 256 14.1 2621 3 US-07-769-996B-36 Sequence 36, Appl
19 256 14.1 2621 3 US-07-769-996C-36 Sequence 36, Appl
20 256 14.1 2621 4 US-08-099-706-36 Sequence 36, Appl
21 256 14.1 3010 3 US-07-635-451-2 Sequence 2, Appl
22 256 14.1 3010 3 US-07-635-451-12 Sequence 12, Appl
23 256 14.1 3010 3 US-07-769-996B-2 Sequence 2, Appl
24 256 14.1 3010 3 US-07-769-996B-14 Sequence 14, Appl
25 256 14.1 3010 3 US-07-769-996C-2 Sequence 2, Appl
26 256 14.1 3010 3 US-07-769-996C-14 Sequence 14, Appl
27 256 14.1 3010 4 US-08-099-706-2 Sequence 2, Appl
28 256 14.1 3010 4 US-08-099-706-14 Sequence 14, Appl
29 256 14.1 3010 7 US-08-383-682-2 Sequence 2, Appl
30 256 14.1 3010 7 US-08-383-682-12 Sequence 12, Appl
31 255 14.0 2985 28 US-10-259-275-60 Sequence 40, Appl
32 255 14.0 3011 21 US-09-742-659-4 Sequence 4, Appl
33 255 14.0 3011 23 US-09-891-894-3 Sequence 3, Appl
34 255 14.0 3011 27 US-10-184-150-3 Sequence 3, Appl
35 255 14.0 3011 31 US-06-300-971-3 Sequence 2, Appl
36 255 14.0 3012 16 US-09-228-076-2 Sequence 2, Appl
37 255 14.0 3012 24 US-09-917-563-2 Sequence 2, Appl
38 255 14.0 3012 25 US-09-963-433-2 Sequence 2, Appl
39 255 14.0 3012 25 US-09-963-433-2 Sequence 2, Appl
40 254 14.0 859 8 US-09-995-937-2 Sequence 2, Appl
41 254 14.0 859 8 US-08-403-5908B-30 Sequence 30, Appl
42 254 14.0 859 8 US-08-444-112-30 Sequence 30, Appl
43 254 14.0 1771 21 US-09-721-4798-2 Sequence 2, Appl
44 254 14.0 1771 21 US-09-721-4798-4 Sequence 4, Appl
45 254 14.0 1771 21 US-09-721-4798-9 Sequence 9, Appl
46 254 14.0 1771 21 US-09-721-4798-11 Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-09-491-146-52
Sequence 52, Application US/09491146
GENERAL INFORMATION:
APPLICANT: KHUDYAKOV, YURI E.
FIELDS, HOWARD A.
TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING THE SA
NUMBER OF SPONSORS: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: JONES & ASKEW, LLP
STREET: 191 Peachtree Street, N.W., 37th Floor
CITY: Atlanta
STATE: GA
COUNTRY: USA
ZIP: 30303-1769
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/491,146
FILING DATE: 25-Jan-2000
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/921,887
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: WARREN, WILLIAM L.
REGISTRATION NUMBER: 36,714
REFERENCE/DOCKET NUMBER: 03063-0380
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-818-3700
TELEFAX: 404-818-3799
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 352 amino acids
TYPE: amino acid
STRANDEDNESS: Not Relevant
TOPOLOGY: Not Relevant
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: Hepatitis virus
SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-09-491-146-52

Query Match 100.0%; Score 1818; DB 18; Length 352;
Best Local Similarity 100.0%; Pred. No. 3.1e-156;
Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAHIFPLEOGMHLAEQFKOKALRPVAPIDREVLYOEFDMEEAASHLPYIEOGMQLAEQFK 60
DB 1 AAHIFPLEOGMHLAEQFKOKALRPVAPIDREVLYOEFDMEEAASHLPYIEOGMQLAEQFK 60
QY 61 OKALAFASRGNHVSPTHYVESDARPAIIPDREVLYHREFDMEEAFAASRGNHVSPTHY 120
DB 61 OKALAFASRGNHVSPTHYVESDARPAIIPDREVLYHREFDMEEAFAASRGNHVSPTHY 120
QY 121 ESDASQAAPYIEQAOVIAHQFKEKVLAFASRGNHVSPTHYVESDARPAIIPDREVLY 180
DB 121 ESDASQAAPYIEQAOVIAHQFKEKVLAFASRGNHVSPTHYVESDARPAIIPDREVLY 180
QY 181 FDEMEESQHLPIEEOGMMLAEQFKOKALAFASRGNHVAPTHYTESDAPALVPDKEVLY 240
DB 181 FDEMEESQHLPIEEOGMMLAEQFKOKALAFASRGNHVAPTHYTESDAPALVPDKEVLY 240
QY 241 QOYDEMEEAFAASRGNHVAPTHYVESDASASLPYMDETRAIAGQFKEKVLAFASRGNHVS 300
DB 241 QOYDEMEEAFAASRGNHVAPTHYVESDASASLPYMDETRAIAGQFKEKVLAFASRGNHVS 300
QY 301 PRHYVESEPQVVVTPDKETILYEAFFDEMEASKALIEEGQMAEMLSKIQ 352
DB 301 PRHYVESEPQVVVTPDKETILYEAFFDEMEASKALIEEGQMAEMLSKIQ 352

RESULT 2
US-09-491-146a-52
Sequence 52, Application US/09491146A
GENERAL INFORMATION:
APPLICANT: The Government of the United States of America, as represented by
APPLICANT: The Secretary, Department of Health and Human Services,
APPLICANT: C/o Centers for Disease Control and Prevention
APPLICANT: Khudyakov, Yuri
APPLICANT: Fields, Howard
TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION ENDONUCLEASE ASSISTED
FILE REFERENCE: 14114.0344U2
CURRENT APPLICATION NUMBER: US/09/491,146A
CURRENT FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: 08/921,887
PRIOR FILING DATE: 25-08-97
NUMBER OF SEQ ID NOS: 55
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 52

LENGTH: 352
TYPE: PRP
ORGANISM: Hepatitis C virus
US-09-491-146a-52

Query Match 100.0%; Score 1818; DB 18; Length 352;
Best Local Similarity 100.0%; Pred. No. 3.1e-156;
Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAHIFPLEOGMHLAEQFKOKALRPVAPIDREVLYOEFDMEEAASHLPYIEOGMQLAEQFK 60
DB 1 AAHIFPLEOGMHLAEQFKOKALRPVAPIDREVLYOEFDMEEAASHLPYIEOGMQLAEQFK 60
QY 61 OKALAFASRGNHVSPTHYVESDARPAIIPDREVLYHREFDMEEAFAASRGNHVSPTHY 120
DB 61 OKALAFASRGNHVSPTHYVESDARPAIIPDREVLYHREFDMEEAFAASRGNHVSPTHY 120
QY 121 ESDASQAAPYIEQAOVIAHQFKEKVLAFASRGNHVSPTHYVESDARPAIIPDREVLY 180
DB 121 ESDASQAAPYIEQAOVIAHQFKEKVLAFASRGNHVSPTHYVESDARPAIIPDREVLY 180
QY 181 FDEMEESQHLPIEEOGMMLAEQFKOKALAFASRGNHVAPTHYTESDAPALVPDKEVLY 240
DB 181 FDEMEESQHLPIEEOGMMLAEQFKOKALAFASRGNHVAPTHYTESDAPALVPDKEVLY 240
QY 241 QOYDEMEEAFAASRGNHVAPTHYVESDASASLPYMDETRAIAGQFKEKVLAFASRGNHVS 300
DB 241 QOYDEMEEAFAASRGNHVAPTHYVESDASASLPYMDETRAIAGQFKEKVLAFASRGNHVS 300
QY 301 PRHYVESEPQVVVTPDKETILYEAFFDEMEASKALIEEGQMAEMLSKIQ 352
DB 301 PRHYVESEPQVVVTPDKETILYEAFFDEMEASKALIEEGQMAEMLSKIQ 352

RESULT 3
US-09-881-239-5
Sequence 5, Application US/09881239
GENERAL INFORMATION:
APPLICANT: CHEN, David Y.
APPLICANT: ARCANDEL, Phillip
APPLICANT: TANDESKE, Laura
APPLICANT: GEORGE-NASCIMENTO, Carlos
APPLICANT: COIT, Doris
APPLICANT: MEDINA-SELBY, Angelica
TITLE OF INVENTION: HCV ANTIGEN/ANTIBODY COMBINATION ASSAY
FILE REFERENCE: 2302-16073 / PP16073.003
CURRENT APPLICATION NUMBER: US/09/881,239
CURRENT FILING DATE: 2001-06-14
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 829
TYPE: PRP
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: MEPA 12
US-09-881-239-5

Query Match 28.2%; Score 512.5; DB 23; Length 829;
Best Local Similarity 62.8%; Pred. No. 6.8e-37;
Matches 118; Conservative 6; Mismatches 15; Indels 49; Gaps 6;

QY 167 KPAILPDBEVLYREFDEMEESQHLPIEEOGMMLAEQFKOKALAFASRGNHVAPTHY 225
DB 167 KPAILPDBEVLYREFDEMEESQHLPIEEOGMMLAEQFKOKALAFASRGNHVAPTHY 225
QY 397 KPAILPDBEVLYREFDEMEESQHLPIEEOGMMLAEQFKOKALAFASRGNHVAPTHY 444
DB 397 KPAILPDBEVLYREFDEMEESQHLPIEEOGMMLAEQFKOKALAFASRGNHVAPTHY 444
QY 226 SAKALVPDKEVLYQOYDEMEEAFAASRGNHVAPTHYVESDASASLPYMDETRAIAGQF 285
DB 226 SAKALVPDKEVLYQOYDEMEEAFAASRGNHVAPTHYVESDASASLPYMDETRAIAGQF 285
QY 445 --GKPAIVDPKEVLYQOYDEMEEAFAASRGNHVAPTHYVESDASASLPYMDETRAIAGQF 483
DB 445 --GKPAIVDPKEVLYQOYDEMEEAFAASRGNHVAPTHYVESDASASLPYMDETRAIAGQF 483
QY 286 KEKVLAFASRGNHVSPTHYVESDARPAIIPDREVLYHREFDMEEAFAASRGNHVSPTHY 344
DB 286 KEKVLAFASRGNHVSPTHYVESDARPAIIPDREVLYHREFDMEEAFAASRGNHVSPTHY 344
QY 484 KEKVLGLI-----DNDQVVVTPDKETILYEAFFDEMEASKALIEEGQMA 529
DB 484 KEKVLGLI-----DNDQVVVTPDKETILYEAFFDEMEASKALIEEGQMA 529

QY 345 EMLKSKIQ 352
Db 530 EMLKSKIQ 537

RESULT 4
US-09-791-537-91118
; Sequence 9118, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9118
; LENGTH: 839
; TYPE: PRT
; ORGANISM: synthetic construct
US-09-791-537-91118

Query Match 28.2%; Score 512.5; DB 22; Length 839;
Best Local Similarity 62.8%; Pred. No. 6.9e-37;
Matches 118; Conservative 6; Mismatches 15; Indels 49; Gaps 6;

QY 167 KPAIIPDREVLVREFPEMEE-SCHLPYIEQGMMLAEQFOKALAFASRGNHVAPTHVTE 225
Db 489 KPAIIPDREVLVREFPEMECSQHLPIYIEQGMMLAEQFOKALGL-SRG----- 536
QY 226 SPAKPLVPDKKVLVYQYDEMEEFASRGNHVAPTHVVESSASLPYMDETRAIAGOF 285
Db 537 --GKPAIVDPKVLVYQYDEME-----CSQAAPYIEQAQVIHAQF 575
QY 286 KEKVLAFASRGNHVSPRYVPESEPOVVTTPDKELIYEAFDEME-ASKAALIEEGORMA 344
Db 576 KEKVLGLI-----DNDQVVVTPDKELIYEAFDEMECSQAALIEEGORMA 621
QY 345 EMLKSKIQ 352
Db 622 EMLKSKIQ 629

RESULT 5
US-09-881-654-4
; Sequence 4, Application US/09881654
; GENERAL INFORMATION:
; APPLICANT: CHEN, David Y.
; APPLICANT: ARANGEL, Phillip
; APPLICANT: TANDESKE, Laura
; APPLICANT: GEORGE-NASCIENTO, Carlos
; APPLICANT: COLT, Doris
; APPLICANT: MEDINA-SELBY, Angelica
; TITLE OF INVENTION: IMMUNOSSAYS FOR ANTI-HCV ANTIBODIES
; FILE REFERENCE: 2302-17039 / PPI7039,002
; CURRENT APPLICATION NUMBER: US/09/881,654
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: 60/212,082
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 60/280,811
; PRIOR FILING DATE: 2001-04-02
; APPLICATION NUMBER: 60/280,867
; FILING DATE: 2001-04-02
; NOS: 7
; 2.0

FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MEPA 7.1
US-09-881-654-4

Query Match 28.2%; Score 512.5; DB 23; Length 1099;
Best Local Similarity 62.8%; Pred. No. 1.1e-36;
Matches 118; Conservative 6; Mismatches 15; Indels 49; Gaps 6;

QY 167 KPAIIPDREVLVREFPEMEE-SCHLPYIEQGMMLAEQFOKALAFASRGNHVAPTHVTE 225
Db 701 KPAIIPDREVLVREFPEMECSQHLPIYIEQGMMLAEQFOKALGL-SRG----- 748
QY 226 SPAKPLVPDKKVLVYQYDEMEEFASRGNHVAPTHVVESSASLPYMDETRAIAGOF 285
Db 749 --GKPAIVDPKVLVYQYDEME-----CSQAAPYIEQAQVIHAQF 787
QY 286 KEKVLAFASRGNHVSPRYVPESEPOVVTTPDKELIYEAFDEME-ASKAALIEEGORMA 344
Db 788 KEKVLGLI-----DNDQVVVTPDKELIYEAFDEMECSQAALIEEGORMA 833
QY 345 EMLKSKIQ 352
Db 834 EMLKSKIQ 841

RESULT 6
US-08-417-478-1
; Sequence 1, Application US/08417478
; GENERAL INFORMATION:
; APPLICANT: CASEY, JAMES M.
; APPLICANT: BODE, SUZANNE L.
; APPLICANT: ZECK, BILLY J.
; APPLICANT: YAMAGUCHI, JULIE
; APPLICANT: FRAIL, DONALD E.
; APPLICANT: DESAI, SURESH M.
; APPLICANT: DEVARE, SUSHIL G.
; TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/417,478
; FILING DATE:
; CLASSIFICATION: 435
; APPLICATION DATA:
; APPLICATION NUMBER: US/08/144,099
; FILING DATE:
; APPLICATION NUMBER: US 07/830,024
; FILING DATE: 01-JAN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMBSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5131, US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-9556
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3011 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

Mon Jul 28 11:16:34 2003

Sequence 10, Application US/07635451
GENERAL INFORMATION:
APPLICANT: OKAYAMA, Hiroto
APPLICANT: FUKU, Isao
APPLICANT: MORI, Chisato
APPLICANT: TAKAMIZAWA, Akahisa
APPLICANT: YOSHIDA, Iwao
TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Armstrong, Nikaido, Marmelstein, Kubovcik, &
Murray
STREET: 1725 K St. N.W. Suite 1000
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/635,451
FILING DATE: 19901228
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-167466
FILING DATE: 25-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-230921
FILING DATE: 31-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-305605
FILING DATE: 09-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: O'Toole, J. Herbert
REGISTRATION NUMBER: 31,404
REFERENCE/DOCKET NUMBER: 900703
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 659-2930
TELEFAX: (202) 887-0357
TELEX: 440142
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2013 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-635-451-10

Query Match 14.1%; Score 256; DB 3; Length 2013;
Best Local Similarity 31.1%; Pred. No. 5.8e-13;
Matches 83; Conservative 31; Mismatches 75; Indels 78; Gaps 10;
DB 23 RPAVDPREVLYOEEDMEE-ASHLPYIEGQMLAEQFKOKALAF---ASRONHVSPTHY 78
1691 RPAVDPRELLYOEEDMEECASHLPYIEGQMLAEQFKOKALGLLOTATKQAEA----- 1745
QY 79 VPESDARPAIIPREVLRHREFDEMEAFASR-GNHVSPAHYVPESDAQAPYIEQ----- 133
DB 1746 ----AAP-----VVEKRALETFWAKHMMNFIISGIQYLAGLSTLPGNPAIASLMAF 1793
QV 134 -AOVIAHOFKEKYLAF-----ASRGNDSPTHYVPESDAQAPAI----- 170
794 TAITSLTOSTLFLNIIIGWVAALAPSAASATVAGACIAGAAVGSIGLGKVLVDILA 1853
-IPREVLYREFDEMEESQHL--PYIEGQML-----AE 201
--MPSTEDLVNLLPALISPGALVGVCAILRRHVGPGEAV 1913

DB 1914 QMMNRLIAFASRGHVSPTHYVPESDA 1940

RESULT 10
US-07-769-996B-12
Sequence 12, Application US/07769996B
GENERAL INFORMATION:
APPLICANT: OKAYAMA, Hiroto
APPLICANT: FUKU, Isao
APPLICANT: MORI, Chisato
APPLICANT: TAKAMIZAWA, Akahisa
APPLICANT: YOSHIDA, Iwao
TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Armstrong, Nikaido, Marmelstein
Kubovcik & Murray
STREET: 1725 K St. N.W. Suite 1000
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/769,996B
FILING DATE: 19911002
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-167466
FILING DATE: 25-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-230921
FILING DATE: 31-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-305605
FILING DATE: 09-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/635,451
FILING DATE: 28-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: O'Toole, J. Herbert
REGISTRATION NUMBER: 31,404
REFERENCE/DOCKET NUMBER: 900703B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 659-2930
TELEFAX: (202) 887-0357
TELEX: 440142
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2013 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-769-996B-12
Query Match 14.1%; Score 256; DB 3; Length 2013;
Best Local Similarity 31.1%; Pred. No. 5.8e-13;
Matches 83; Conservative 31; Mismatches 75; Indels 78; Gaps 10;
DB 23 RPAVDPREVLYOEEDMEE-ASHLPYIEGQMLAEQFKOKALAF---ASRONHVSPTHY 78
1691 RPAVDPRELLYOEEDMEECASHLPYIEGQMLAEQFKOKALGLLOTATKQAEA----- 1745
QY 79 VPESDARPAIIPREVLRHREFDEMEAFASR-GNHVSPAHYVPESDAQAPYIEQ----- 133
DB 1746 ----AAP-----VVEKRALETFWAKHMMNFIISGIQYLAGLSTLPGNPAIASLMAF 1793

OY 134 -AQTIAHQFKEKVLAF-----ASRGNDSPTHVYVESDAKPAI-----170
DB 1794 TASITSPITTOSTLFTNLIGGVAAQLAPPSAASAFVAGAGAAVSGIGLKVLDILA 1853
OY 171 -----IPDREVLVREPFDEMEESQHL--PYIEQGMML-----AE 201
DB 1854 GYGAGVAGALVAFKVMGSEMPSTEDLVNLLPALISPGALVGVCAAILRRHVGPGECAV 1913
OY 202 QFKOKALAFASRGNHVAPTHVYTESDA 228
DB 1914 QMNRRLIAFASRGNHVSPTHVYVESDA 1940

RESULT 11
US-07-769-996C-12
Sequence 12, Application US/07769996C
GENERAL INFORMATION:
APPLICANT: OKAYAMA, Hiroto
APPLICANT: FUKU, Isao
APPLICANT: MORI, Chisato
APPLICANT: TAKAMIZAWA, Akahisa
APPLICANT: YOSHIDA, Iwao
TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Armstrong, Westerman, Hattori, Mclelland
STREET: 1725 K St. N.W. Suite 1000
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/769,996C
FILING DATE: 02-OCT-1991
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-167466
FILING DATE: 25-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-230921
FILING DATE: 31-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-305605
FILING DATE: 09-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/635,451
FILING DATE: 28-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: Mclelland, Le-Nhung
REGISTRATION NUMBER: 31,541
REFERENCE/DOCKET NUMBER: 900703B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 659-2930
TELEFAX: (202) 887-0357
TELEX: 440142
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2013 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-769-996C-12

Query Match 14.1%; Score 256; DB 3; Length 2013;
Best Local Similarity 31.1%; Pred. No. 5.8e-13;
Matches 83; Conservative 31; Mismatches 75; Indels 78; Gaps 10;

OY 23 RPAVIPPDEVLVYOEPEDEME-ASHLPYIEQGMLOAFQOKALAF-----ASRGNDSPTHV 78
DB 1691 RPAVIPPDEVLVYOEPEDEMECAHLPYIEQGMLOAFQOKALGLLQITKOE-A-----1745
OY 79 VPESDARPAIIPDREVLVREPFDEMEEAFAFR-GNHVSPAHVYVESDASQAAPYIEO-----133
DB 1746 -----AAP-----VSESKRALETFMKHMNFIISGIQYLAGLSTLPONPAIASLMAF 1793
OY 134 -AQTIAHQFKEKVLAF-----ASRGNDSPTHVYVESDAKPAI-----170
DB 1794 TASITSPITTOSTLFTNLIGGVAAQLAPPSAASAFVAGAGAAVSGIGLKVLDILA 1853
OY 171 -----IPDREVLVREPFDEMEESQHL--PYIEQGMML-----AE 201
DB 1854 GYGAGVAGALVAFKVMGSEMPSTEDLVNLLPALISPGALVGVCAAILRRHVGPGECAV 1913
OY 202 QFKOKALAFASRGNHVAPTHVYTESDA 228
DB 1914 QMNRRLIAFASRGNHVSPTHVYVESDA 1940

RESULT 12
US-08-099-706-12
Sequence 12, Application US/08099706
GENERAL INFORMATION:
APPLICANT: OKAYAMA, Hiroto
APPLICANT: FUKU, Isao
APPLICANT: MORI, Chisato
APPLICANT: TAKAMIZAWA, Akahisa
APPLICANT: YOSHIDA, Iwao
TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Armstrong, Nikaido, Marmelstein
ADDRESSEE: Kubovcik & Murray
STREET: 1725 K St. N.W. Suite 1000
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/099,706
FILING DATE: 30-JUL-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-167466
FILING DATE: 25-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-230921
FILING DATE: 31-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-305605
FILING DATE: 09-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/635,451
FILING DATE: 28-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/769,996B
FILING DATE: 02-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: O'Toole, J. Herbert
REGISTRATION NUMBER: 31,404
REFERENCE/DOCKET NUMBER: 900703B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 659-2930
TELEFAX: (202) 887-0357

TELEX: 440142
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2013 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-099-706-12

Query Match 14.1%; Score 256; DB 4; Length 2013;
Best Local Similarity 31.1%; Pred. No. 5.8e-13;
Matches 83; Conservative 31; Mismatches 75; Indels 78; Gaps 10;

QY 23 RPVAVDPREVLYGFEDMEF--ASHLPYIEQGMQLAEQKOKALAF--ASRQNHVSPTHY 78
DB 1691 RPAIVDPRELLYGFEDMEECASHLPYIEQGMQLAEQKOKALGLQTATKQAE----- 1745
QY 79 VPESDARPAIIPREVLYHREFDEMEAFASR--GNHVSFAHYVPESDASQAPYIEQ---- 133
DB 1746 -----AAP-----VVESEKRALETFWAKHMMNFISGLOYLAGLSTLPGNPALIASLMAF 1793
QY 134 -AQVIAHQFEKVLAF-----ASRGNHDSPTHYVPESDAKPAI----- 170
DB 1794 TASITSPLTQSTLFLNIGGVAAQLAPPSASAFVAGAGIAGAAVSGIGKVLVDILA 1853
QY 171 -----IPREVLYREFDEMEESQHL--PYIEQGMML-----AE 201
DB 1854 GYGAGVAGALVAFKVMGSEMPSTEDLVNLLPALISGALVGVCAALIRRHVGPEGCAY 1913
QY 202 QFKOKALAFASRGNHVAPTHYVTESDA 228
DB 1914 QMMNRLIAFASRGNHVSPTHYVTESDA 1940

RESULT 13
US-08-383-682-10
Sequence 10, Application US/08383682
GENERAL INFORMATION:
APPLICANT: OKAYAMA, Hiroko
APPLICANT: FUKU, Isao
APPLICANT: MORI, Chisato
APPLICANT: TAKAMIZAWA, Akahisa
APPLICANT: YOSHIDA, Iwao
TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS PARTICLES
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Armstrong, Nikolaide, Marmelstein, Kubovcik, &
STREET: 1725 K St. N.W. Suite 1000
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/383,682
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/720,961
FILING DATE: 25-JUN-1991
APPLICATION NUMBER: JP 2-167466
DATE: 25-JUN-1990
ON DATA: JP 2-230921

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/635,451
FILING DATE: 28-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: O'Toole, J. Herbert
REGISTRATION NUMBER: 31,404
REFERENCE/DOCKET NUMBER: 900703A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 659-2930
TELEFAX: (202) 887-0357
TELEX: 440142
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2013 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-383-682-10

Query Match 14.1%; Score 256; DB 7; Length 2013;
Best Local Similarity 31.1%; Pred. No. 5.8e-13;
Matches 83; Conservative 31; Mismatches 75; Indels 78; Gaps 10;

QY 23 RPVAVDPREVLYGFEDMEF--ASHLPYIEQGMQLAEQKOKALAF--ASRQNHVSPTHY 78
DB 1691 RPAIVDPRELLYGFEDMEECASHLPYIEQGMQLAEQKOKALGLQTATKQAE----- 1745
QY 79 VPESDARPAIIPREVLYHREFDEMEAFASR--GNHVSFAHYVPESDASQAPYIEQ---- 133
DB 1746 -----AAP-----VVESEKRALETFWAKHMMNFISGLOYLAGLSTLPGNPALIASLMAF 1793
QY 134 -AQVIAHQFEKVLAF-----ASRGNHDSPTHYVPESDAKPAI----- 170
DB 1794 TASITSPLTQSTLFLNIGGVAAQLAPPSASAFVAGAGIAGAAVSGIGKVLVDILA 1853
QY 171 -----IPREVLYREFDEMEESQHL--PYIEQGMML-----AE 201
DB 1854 GYGAGVAGALVAFKVMGSEMPSTEDLVNLLPALISGALVGVCAALIRRHVGPEGCAY 1913
QY 202 QFKOKALAFASRGNHVAPTHYVTESDA 228
DB 1914 QMMNRLIAFASRGNHVSPTHYVTESDA 1940

RESULT 14
US-10-085-476-2
Sequence 2, Application US/10085476
GENERAL INFORMATION:
APPLICANT: De Francesco, Raffaele
APPLICANT: Tomel, Licia
APPLICANT: Behrens, Sven-Erik
TITLE OF INVENTION: METHOD FOR REPRODUCING IN VITRO THE
TITLE OF INVENTION: RNA-DEPENDENT RNA POLYMERASE AND TERMINAL NUCLEOTIDYL
TRANSFERASE ACTIVITIES ENCODED BY HEPATITIS C VIRUS (HCV)
FILE REFERENCE: IT0002PCA
CURRENT APPLICATION NUMBER: US/10/085,476
PRIOR FILING DATE: 2002-02-27
PRIOR APPLICATION NUMBER: 08/952,981
PRIOR FILING DATE: 1998-03-23
PRIOR APPLICATION NUMBER: PCT/IT96/00106
PRIOR FILING DATE: 1996-05-24
PRIOR APPLICATION NUMBER: RM95A000343
PRIOR FILING DATE: 1995-05-25
NUMBER OF SEQ ID NOS: 14
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 2201
TYPE: PRT
ORGANISM: CDNA clone pCD (38-9.4)
US-10-085-476-2

Query Match 14.1%; Score 256; DB 26; Length 2201;
Best Local Similarity 31.1%; Pred. No. 6.7e-13;

Mon Jul 28 11:16:34

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 13:06:21 / Search time 97 Seconds

(without alignments)
936.438 Million cell updates/sec

Title: US-09-491-146a-52

Perfect score: 1918
Sequence: 1 AAH1PYLEQGHLEDFKOK.....KAALIEGQMAENLKSIIQ 352

Scoring table:

BLOSUM62
Gapop 10.0, Gapext 0.5

Sched: 830525 seqs, 258052604 residues

1 number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	268	14.7	3008	12 039929	039929 hepatitis c
2	268	14.7	3010	12 093363	093363 hepatitis c
3	268	14.7	3010	12 081825	081825 hepatitis c
4	268	14.7	3010	12 068285	068285 hepatitis c
5	263	14.5	3010	12 0901X1	0901X1 hepatitis c
6	263	14.5	3010	12 068949	068949 hepatitis c
7	261	14.4	3010	12 09DTE2	09DTE2 hepatitis c
8	261	14.4	3010	12 0933H9	0933H9 hepatitis c
9	261	14.4	3010	12 0901X2	0901X2 hepatitis c
10	260	14.3	3010	12 093016	093016 hepatitis c
11	260	14.3	3010	12 0933G7	0933G7 hepatitis c
12	260	14.3	3010	12 09DTE5	09DTE5 hepatitis c
13	259	14.2	3008	12 0933F4	0933F4 hepatitis c
14	259	14.2	3010	12 068788	068788 hepatitis c
15	259	14.2	3010	12 0901X3	0901X3 hepatitis c
16	259	14.2	3010	12 090P61	090P61 hepatitis c

17	259	14.2	3010	12 0901X4	0901X4 hepatitis c
18	259	14.2	3010	12 093077	093077 hepatitis c
19	259	14.2	3010	12 09DTE8	09DTE8 hepatitis c
20	259	14.2	3010	12 09DTE6	09DTE6 hepatitis c
21	259	14.2	3010	12 081757	081757 hepatitis c
22	259	14.2	3011	12 003463	003463 hepatitis c
23	259	14.2	3013	12 0901X9	0901X9 hepatitis c
24	258	14.2	3010	12 P88803	P88803 hepatitis c
25	258	14.2	3010	12 0933H8	0933H8 hepatitis c
26	258	14.2	3010	12 0901Z0	0901Z0 hepatitis c
27	258	14.2	3010	12 09DTE1	09DTE1 hepatitis c
28	258	14.2	3010	12 0933H5	0933H5 hepatitis c
29	258	14.2	3010	12 0933F9	0933F9 hepatitis c
30	258	14.2	3010	12 0933G6	0933G6 hepatitis c
31	258	14.2	3010	12 0901Y9	0901Y9 hepatitis c
32	258	14.2	3010	12 P90191	P90191 hepatitis c
33	258	14.2	3011	12 0933F5	0933F5 hepatitis c
34	258	14.2	3013	12 0901Y0	0901Y0 hepatitis c
35	257	14.1	3010	12 0901Y1	0901Y1 hepatitis c
36	257	14.1	3010	12 09DTE9	09DTE9 hepatitis c
37	257	14.1	3010	12 0901Y2	0901Y2 hepatitis c
38	257	14.1	3010	12 0901X8	0901X8 hepatitis c
39	257	14.1	3010	12 0901X7	0901X7 hepatitis c
40	257	14.1	3011	12 0901X8	0901X8 hepatitis c
41	257	14.1	3011	12 09DTE3	09DTE3 hepatitis c
42	256	14.1	2284	12 081817	081817 hepatitis c
43	256	14.1	3010	12 0901Y3	0901Y3 hepatitis c
44	256	14.1	3010	12 081760	081760 hepatitis c
45	256	14.1	3010	12 P89966	P89966 hepatitis c

ALIGNMENTS

RESULT 1

039929 PRELIMINARY; PRT; 3008 AA.

AC 039929;
RC 01-JAN-1998 (TREMURel. 05, Created)
DT 01-JAN-1998 (TREMURel. 23, Last sequence update)
DE HCV polypeptide (Genome polypeptide).
OS Hepatitis C virus type 4a.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus;
OX NCBI_Taxid=31653;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ED43 genotype 4A; Pubmed=9191927;
RX MEDLINE=97335261; Pubmed=9191927;
RA Chamberlain R.W., Adams N., Saeed A.A., Simmonds P., Elliott R.M.;
RT "Complete nucleotide sequence of a type 4 hepatitis C virus variant,
the predominant genotype in the Middle East."
J. Gen. Virol. 78:1341-1347(1997).
RL -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
PROTEIN C AND RNA (BY SIMILARITY).
EMBL; Y11604; CA72338.1; -;
DR HSSP; P27958; 1HE1.
DR MEROPS; S29.001; -;
DR MEROPS; U39.001; -;
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRp.

RN [1]
 RP Hepatitis C virus.",
 RA Submitted (JUN-1993) to the EMBL/GenBank/DDBJ databases.
 RL -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MNA (BY SIMILARITY).
 EMBL; M96162; AAA45721.1; -.
 DR HSP, P26663; LUXP.
 DR InterPro: IPR001410, DEAD.
 DR InterPro: IPR002522, HCV capsid.
 DR InterPro: IPR002521, HCV core.
 DR InterPro: IPR002519, HCV env.
 DR InterPro: IPR002531, HCV NS1.
 DR InterPro: IPR002518, HCV NS2.
 DR InterPro: IPR004109, HCV NS3.
 DR InterPro: IPR000745, HCV NS4a.
 DR InterPro: IPR001490, HCV NS4b.
 DR InterPro: IPR002868, HCV NS5a.
 DR InterPro: IPR002166, HCV RdRP.
 DR InterPro: IPR007095, RNA pol_PS.
 DR InterPro: IPR007094, RNA pol_PsVtr.
 DR Pfam: PF01543, HCV capsid; 1.
 DR Pfam: PF01542, HCV core; 1.
 DR Pfam: PF01535, HCV env; 1.
 DR Pfam: PF01560, HCV NS1; 1.
 DR Pfam: PF01538, HCV NS2; 1.
 DR Pfam: PF02907, HCV NS3; 1.
 DR Pfam: PF01006, HCV NS4a; 1.
 DR Pfam: PF01001, HCV NS4b; 1.
 DR Pfam: PF01506, HCV NS5a; 1.
 DR Pfam: PF00998, Viral_RdRP; 1.
 DR ProDom: PD186062, HCV NS1; 1.
 DR SMART: SM00487, DEXDC; 1.
 DR PROSITE: PS50507, RDRP POSITIVE; 1.
 DR PROSITE: PS50521, RDRP VIRAL; 1.
 KM Coat protein, Envelope protein, Glycoprotein, Nonstructural protein,
 KW Polyprotein, RNA-directed RNA polymerase, Transferase, Transmembrane,
 QO SEQUENCE 3010 AA; 326924 MW; FE997D54EE05142B CRC64;

[illegible]

ID	Q68285	PRELIMINARY;	PRT; 3010 AA.
AC	068285;		
DT	01-NOV-1996	(TREMBLrel. 01, Created)	
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)	
DT	01-MAR-2003	(TREMBLrel. 23, Last annotation update)	
DE	Genome polypeptide.		
OS	Hepatitis C virus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.		
CC	NCBI_TaxID=11103;		
RP	[1]		
RA	SEQUENCE FROM N.A.		
RT	Cho J., Park Y., Lee Y., Yang J., Ryu W.;		
RL	"Molecular cloning of Hepatitis C virus genes from chronic patients in Korea";		
RC	Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.		
CC	-1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MNA (BY SIMILARITY).		
CC	EMBL; U16362; AAA52748.1; --		
CC	HSSP; P26663; LUXP.		
DR	InterPro; IPR001410; DEAD.		
DR	InterPro; IPR002522; HCV_capsid.		
DR	InterPro; IPR002521; HCV_core.		
DR	InterPro; IPR002519; HCV_env.		
DR	InterPro; IPR002531; HCV_NSI.		
DR	InterPro; IPR002518; HCV_NS2.		
DR	InterPro; IPR004109; HCV_NS3.		
DR	InterPro; IPR000745; HCV_NS4a.		
DR	InterPro; IPR001490; HCV_NS4b.		
DR	InterPro; IPR002868; HCV_NS5a.		
DR	InterPro; IPR002166; HCV_RdRP.		
DR	InterPro; IPR007095; RNA_pol_DS_PS.		
DR	InterPro; IPR007094; RNA_pol_PSVir.		
DR	pfam; PF01543; HCV_capsid; 1.		
DR	pfam; PF01542; HCV_core; 1.		
DR	pfam; PF01539; HCV_env; 1.		
DR	pfam; PF01560; HCV_NSI; 1.		
DR	pfam; PF01538; HCV_NS2; 1.		
DR	pfam; PF02907; HCV_NS3; 1.		
DR	pfam; PF01006; HCV_NS4a; 1.		
DR	pfam; PF01001; HCV_NS4b; 1.		
DR	pfam; PF01506; HCV_NS5a; 1.		
DR	pfam; PF00998; Viral_RdRP; 1.		
DR	ProDom; PD186062; HCV_NSI; 1.		
DR	SMART; SM00487; DEXDC; 1.		
DR	PROSITE; PS50507; RDRP_POSITIVE; 1.		
DR	PROSITE; PS50521; RDRP_VIRAL; 1.		
KW	Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.		
SC	SEQUENCE 3010 AA; 326914 MW; 298306FC58B8BC9E CRC64;		
Query Match	14.7%; Score 268; DB 12; Length 3010;		
Best Local Similarity	30.7%; Pred. No. 1.7e-10;		
Matches 102; Conservative 26; Mismatches 88; Indels 116; Gaps 14;			

Query Match	14.7%	Score 268;	DB 12;	Length 3010;
Best Local Similarity	30.7%	Pred. No. 1.7e-10;		
Matches	102;	Conservative	26;	Mismatches 88; Indels 116; Gaps 14;
QY	23	RPAAVTPREVLYQGFDEME--ASHLPYIEQGMQALAEQFKOKALF--ASFQNHVSPHY	78	
DB	1691	KPAIIPPREVLVYQGFDEMEBCASHLPYEQGMQALAEQFKOKALGILQATATQAE-	1745	
QY	79	VPESDAPAIIPREVLRFEDEMEELFASR--GNHVSFAHYVPESDASQAAPYIEQAQVI	137	
DB	1746	----AAD-----VVESKRALETTPAKMMNFISTQIQLAGLSTLPGNPALRSP---	1790	
QY	138	AHQEKVLAFAASRGNDHSPHYVPESDAKPAIIPREVLYREFDEMESH--LPYIEQG	196	
DB	1791	-----MART-----SITSPL-----TTQHTLLEFNILG	1813	
QY	197	MLLAEOFRQKALAPASRGNHVAPHYHTESPAK-----PALVPDKEVLYQ	242	
DB	1814	GMVAALQALAPSPASAAFGVAGIAGAVALGTIGLGVKLVLDLAGYGAAGVALAFK--INSGE	1877	

QY 243 YDEME-----AFASRG-----NHVAPTHYVESDASASLPYMETRAIAG 283
 Db 1873 MPASAEWVNLPLPALSGALVGVCAIIRRHVGPGEAV-----1913

QY 284 QFKEKVLAFASRGNHVSPRYVPESEPOVVT 315
 Db 1914 QMMNRLIAFASRGNHVSPRYVPESEPAARVT 1945

RESULT 5
 ID 090IX1 PRELIMINARY; PRT; 3010 AA.
 AC 090IX1;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NCBI_TaxID=11103;
 [1]
 SEQUENCE FROM N.A.
 RC STRAIN=MD10-2;
 RX MEDLINE=20013325; Pubmed=10544098;
 RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S.Y., Miyasaka Y.,
 RA Tazawa J.I., Izumi N., Marumo F., Sato C.;
 RT "Time-related changes in full-length hepatitis C virus and hepatitis
 RT activity."
 RL Virology 263:244-253(1999).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MD10-2;
 RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S., Miyasaka Y.,
 RA Sakamoto N., Fukuma T., Tazawa J., Izumi N., Marumo F., Sato C.;
 RT Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT. THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 CC EMBL: AF165064; AAD56199.1; -
 DR HSP; P26663; IJXP.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4.
 DR InterPro: IPR001490; HCV_NS4B.
 DR InterPro: IPR002868; HCV_NS5A.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid.1.
 DR Pfam: PF01542; HCV_core.1.
 DR Pfam: PF01539; HCV_env.1.
 DR Pfam: PF01560; HCV_NS1.1.
 DR Pfam: PF01538; HCV_NS2.1.
 DR Pfam: PF02907; HCV_NS3.1.
 DR Pfam: PF01006; HCV_NS4.1.
 DR Pfam: PF01001; HCV_NS4B.1.
 DR Pfam: PF01506; HCV_NS5A.1.
 DR Pfam: PF00998; Viral_RdRp.1.
 DR Pfam: PD186062; HCV_NS1.1.
 DR SMART: SM00487; DEXDC.1.
 DR PROSITE: PSS0507; RDRP_POSITIVE.1.
 DR PROSITE: PSS0521; RDRP_VIRAL.1.
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
 SO SEQUENCE 3010 AA; 326902 MW; 0423385981E71EDD CMC64;

Query Match 14.5%; Score 263; DB 12; Length 3010;
 Best Local Similarity 29.7%; Pred. No. 3.9e-10;
 Matches 94; Conservative 30; Mismatches 107; Indels 86; Gaps 11;

QY 23 RPAVIDRREVLVQEFDEME-ASHLPYIQGQALAEQFOKALAF---ASRONHVSPTHY 78
 Db 1691 RPAIIRDRREVLVREFDEMECSHLPYIQGQALAEQFOKALGGLQITKQ-----1742

QY 79 VPESDAPPAIIPDREVLAREFDEMEAFASR-GNHVSPAHVVPESDASQAAPYEQAYI 137
 Db 1743 -----ABAAPV---VVESKQWALAEAFMAKHMNFISGIQYLAGSTLPGNPAI-----1787

QY 138 ANQFKEKVLAFASRGNHSDPTHVPESDAKPAIIPDREVLAREFDEMESQHLPTYEQGM 197
 Db 1788 -----ASLMAFTASVTSPLTQYT-----LLFNILGQWVAQIAPPSASSA 1828

QY 198 MLAEQFKQALAFASRGNHVAPTHVYTESDAKPAIIPDREVLVQYQYDEME-----AFA 251
 Db 1829 FPGAGIAGAAVSGISGLKVLVILAGYGVAGALVAFK-VMSGEVPTEDLVNLLPALI 1887

QY 252 SRG-----NHVAPTHYVESDASASLPYMETRAIACQFKEKVLAFASRGNH 298
 Db 1888 SPGALVGVCAIIRRHVGPGEAV-----QMMNRLIAFASRGNH 1928

QY 299 VSPRYVPESEPOVVT 315
 Db 1929 VSPRYVPESEPAARVT 1945

RESULT 6
 ID 068949 PRELIMINARY; PRT; 3010 AA.
 AC 068949;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE UK1-full protein (Genome polyprotein).
 GN UK1-FULL.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NCBI_TaxID=11103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=93013055; Pubmed=1327977;
 RA Honda M., Kaneko S., Unoura U., Kobayashi K., Murakami S.;
 RT "Sequence comparisons for a hepatitis C virus genome RNA isolated from
 RT a patient with liver cirrhosis."
 RL Gene 120:317-318(1992).
 CC -1- SUBUNIT. THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 CC EMBL: X61596; CAA43793.1; -
 DR HSP; P26663; IJXP.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4.
 DR InterPro: IPR001490; HCV_NS4B.
 DR InterPro: IPR002868; HCV_NS5A.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid.1.
 DR Pfam: PF01542; HCV_core.1.
 DR Pfam: PF01539; HCV_env.1.
 DR Pfam: PF01560; HCV_NS1.1.
 DR Pfam: PF01538; HCV_NS2.1.

```
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; Viral_RdRp; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS50507; RDRP_POSITIVE; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
FT CHAIN 1
FT CHAIN 191
FT CHAIN 383
FT CHAIN 384
FT CHAIN 734
FT CHAIN 1007
FT CHAIN 1616
FT CHAIN 1862
FT CHAIN 2013
FT CHAIN 2014
SEQUENCE 3010 AA; 327178 MW; 4517408FF613ADA CRC64;

Query Match 14.5%; Score 263; DB 12; Length 3010;
Best Local Similarity 32.5%; Pred. No. 3,9e-10;
Matches 87; Conservative 27; Mismatches 74; Indels 80; Gaps 11;

QY 23 RPAVLPDEVLVYOEEDMEEE-ASHLPYIEQGMQLAEQFKOKALAF---ASRQNHVSPTHY 78
DB 1691 RPAIIPDEVLVYOEEDMEECASHLPYIEQGMQLAEQFKOKALGLLOTATSKQAEE----- 1745
QY 79 VPESDAPPAIIPDEVLVHREFDEMEEFASR--GNHVSAPHVVPESDASQAAPYIEQ--- 133
DB 1746 -----AAPVESKQWQL-----EAFWAKHMMNFISGIQYLAGLSTLPQNPALVSLMA 1792
QY 134 --AOVIAHQFKEKVLAF-----ASRGNHDSPTHYVPESDAKPAI----- 170
DB 1793 FTASITSPLTQHTLLEFNLIGWVAQAQLAPPSAASAFVAGAGIAGAAVGSIGLKVLDIL 1852
QY 171 -----IPDEVLVYREFD-EMESQHL-----PYIEQGMQL-----A 200
DB 1853 AGYGAGVAGALVAFKMGSEMPSTEDLVNLLPALISPGALVGVCAILRRHVGPEGGA 1912
QY 201 EQFKOKALAFASRGNHVAFTHYTESDA 228
DB 1913 VQMMNRLLAFASRGNHVSPTHYVPESDA 1940

RESULT 7
Q9JTE2 PRELIMINARY; PRT; 3010 AA.
AC Q9JTE2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_Taxid=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV1169;
RA Takehashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
RA Haraheara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,
RA Mishiro S.;
RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
RT with hepatocellular carcinoma: the 'progression score' revisited.";
RT Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC EMBL; AB049095; BAB18808.1; -.
DR EMBL; P26663; JUXP.
```

```
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRp.
DR InterPro; IPR007095; RNA_pol_DS_P5.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; Viral_RdRp; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00190; CYTOCHROME C; 1.
DR PROSITE; PS50507; RDRP_POSITIVE; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SEQUENCE 3010 AA; 326867 MW; 25BE54B9D7EEA15 CRC64;

Query Match 14.4%; Score 261; DB 12; Length 3010;
Best Local Similarity 32.1%; Pred. No. 5,5e-10;
Matches 86; Conservative 28; Mismatches 74; Indels 80; Gaps 10;

QY 23 RPAVLPDEVLVYOEEDMEEE-ASHLPYIEQGMQLAEQFKOKALAF---ASRQNHVSPTHY 78
DB 1691 RPAIIPDEVLVYOEEDMEECASHLPYIEQGMQLAEQFKOKALGLLOTATQAEV----- 1745
QY 79 VPESDAPPAIIPDEVLVHREFDEMEEFASR--GNHVSAPHVVPESDASQAAPYIEQ--- 133
DB 1746 -----AAPVESKQWQL-----EAFWAKHMMNFISGIQYLAGLSTLPQNPALVSLMA 1792
QY 134 --AOVIAHQFKEKVLAF-----ASRGNHDSPTHYVPESDAKPAI----- 170
DB 1793 FTASITSPLTQHTLLEFNLIGWVAQAQLAPPSAASAFVAGAGIAGAAVGSIGLKVLDIL 1852
QY 171 -----IPDEVLVYREFD-EMESQHL-----PYIEQGMQL-----A 200
DB 1853 AGYGAGVAGALVAFKMGSDPSTEDLVNLLPALISPGALVGVCAILRRHVGPEGGA 1912
QY 201 EQFKOKALAFASRGNHVAFTHYTESDA 228
DB 1913 VQMMNRLLAFASRGNHVSPTHYVPESDA 1940

RESULT 8
Q9J3H9 PRELIMINARY; PRT; 3010 AA.
AC Q9J3H9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_Taxid=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MD13;
```

RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.:
 RT "Characteristics of hepatitis C viral genome associated with disease
 progression.";
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 DR EMBL; AF207754; AAF65944.1; -.
 DR HSP; P26663; IUXP.
 DR InterPro: IPR000345; Cyt_c_heme_bind.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.
 DR PROSITE; PS05057; RDRP_POSITIVE; 1.
 DR PROSITE; PS05052; RDRP_VIRAL; 1.
 DR CoaT protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KM Polypeptide; RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 326984 MW; AFI2CC0E0A8B078 CRC64;

Query Match 14.4%; Score 261; DB 12; Length 3010;
 Best Local Similarity 32.6%; Pred. No. 5.5e-10;
 Matches 87; Conservative 27; Mismatches 75; Indels 78; Gaps 10;

QY 23 RPAVIPDEVLVYREFDEME--ASHLPYIEGQMLAEQFKOKALAF---ASRGNHVSPTHY 78
 |||||
 DB 1691 RPAVIPREVLYREFDEMECHSHLPYIEGQMLAEQFKOKALGLQATKQAE----- 1745

QY 79 VPESDARPAIIPREVLYREFDEMEAFASR--GNHVSAPHYVPESDASQAAPYIEQ---- 133
 |||||
 DB 1746 -----AAPAV-----ESKMRALLETFWAKHMNFISGIYLAGLSTLPENPAIASLMNF 1793

QY 134 -AOYIAHQFKEXKVLAF-----ASRGNHDSPTHYVPESDAKPAI----- 170
 |||||
 DB 1794 TASVTSPLTQNTLLFNILGWAQAQLAPPSAASAFGAGIAGAAGVSGISGLKVLVDIILA 1853

QY 171 -----IPREVLYREFDEMEESQHL--PYIEGQML-----AE 201
 |||||
 DB 1854 GYAGAGAGALVARKVMSGETPSSAEELVNLPAILSPGALVGVCAILRRHYVPGEGAV 1913

QY 202 QFKOKALAFASRGNHVAPTHYVPESDA 228
 |||||
 DB 1914 QMMNRLIAFASRGNHVSPTHYVPESDA 1940

RESULT 9
 Q90IX2 PRELIMINARY; PRT; 3010 AA.
 AC Q90IX2;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAY-2003 (TrEMBLrel. 23, Last annotation update)
 DE Genome polyprotein.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11103;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=MD10-1;
 RX MEDLINE=20013325; PubMed=10544098;
 RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S.Y., Miyasaka Y.,
 RA Tazawa J., Izumi N., Marumo F., Sato C.:
 RT "Time-related changes in full-length hepatitis C virus and hepatitis
 activity.";
 RL Virology 263:244-253 (1999).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=MD10-1;
 RA Nagayama K., Kurosaki M., Enomoto N., Maekawa S., Miyasaka Y.,
 RA Sakamoto N., Fukuma T., Tazawa J., Izumi N., Marumo F., Sato C.:
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA (BY SIMILARITY).
 DR EMBL; AF165063; AAD56198.1; -.
 DR HSP; P26663; IUXP.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR PROSITE; PS05057; RDRP_POSITIVE; 1.
 DR PROSITE; PS05052; RDRP_VIRAL; 1.
 DR CoaT protein; Envelope protein; Glycoprotein; Nonstructural protein;
 KM Polypeptide; RNA-directed RNA polymerase; Transferase; Transmembrane.
 SQ SEQUENCE 3010 AA; 326855 MW; 3806A4AR819ED552 CRC64;

Query Match 14.4%; Score 261; DB 12; Length 3010;
 Best Local Similarity 29.9%; Pred. No. 5.5e-10;
 Matches 95; Conservative 28; Mismatches 107; Indels 88; Gaps 11;

QY 23 RPAVIPDEVLVYREFDEME--ASHLPYIEGQMLAEQFKOKALAF---ASRGNHVSPTHY 78
 |||||
 DB 1691 RPAIIPREVLYREFDEMECHSHLPYIEGQMLAEQFKOKALGLQATKQAE----- 1745

QY 79 VPESDARPAIIPREVLYREFDEMEAFASR--GNHVSAPHYVPESDASQAAPYIEQAV 136
 |||||
 DB 1746 -----AAPVSKWQAL-----EAFWAKHMNFISGIYLAGLSTLPENPAI----- 1787

QY 137 IAHQFKEXKVLAFASRGNHDSPTHYVPESDAKPAIIPREVLYREFDEMEESQHLPYIEQ 196
 |||||

```

DB 1788 -----ASLMFTASTVSTPLTQYT-----LLENILGCVWVAQIAPPSASS 1827
QY 197 MMLAEQFKOKALAFASRGHNVAPTHVTESDAKPALVPDKVELYQOYDEME-----AF 250
DB 1828 AFGVAGIAGAAVGSIGLKVLDIILAGYAGAGALVAFK-VMSGEVSTELVNLPAI 1886
QY 251 ASRG-----NHVAPTHVVESDASASLPYDETRAIAGQFKKYLAFASRCN 297
DB 1887 LSPGALVGVCAALIRRHVGPEGAV-----QNMNRLIAFASRCN 1927
QY 298 HVSFPHVPESEPOVVT 315
DB 1928 HVSPTHVPESDAAARVT 1945

RESULT 10
QY 093016 PRELIMINARY; PRT; 3010 AA.
AC 093016;
I 01-NOV-1998 (TREMBLrel. 08, Created)
L 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV-A;
RA Trowbridge R.;
RL Submitted (Jul-1997) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HCV-A;
RX MEDLINE=9823263; PubMed=9572551;
RA Trowbridge R., Gowers E.J.;
RT "Molecular cloning of an Australian isolate of hepatitis C virus.";
RL Arch. Virol. 143:501-511 (1998)
CC -I- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA (BY SIMILARITY).
DB EMBL: AJ000009; CAA03854.1; -
DR HSP; P26663; IXP.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RdRp.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVIT.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01539; HCV_core; 1.
DR Pfam: PF01560; HCV_env; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02807; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR ProDom: PD0186062; HCV_NS1; 1.
DR SMART; SMD0487; DEXDC; 1.
DR PROSITE; PS50507; RDRP_POSITIVE; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
KV Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;

```

```

KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
FT CHAIN 1 191 CORE PROTEIN.
FT CHAIN 192 383 E1 PROTEIN.
FT CHAIN 384 746 E2 PROTEIN.
FT CHAIN 747 809 P7 PROTEIN.
FT CHAIN 810 1026 NS2 PROTEIN.
FT CHAIN 1027 1657 NS3 PROTEIN.
FT CHAIN 1658 1711 NS4A PROTEIN.
FT CHAIN 1712 1972 NS4B PROTEIN.
FT CHAIN 1973 2419 NS5A PROTEIN.
FT CHAIN 2420 3010 NS5B PROTEIN.
SQ SEQUENCE 3010 AA; 327120 MW; 0E02EDA54A861D CRC64;

Query Match 14.3%; Score 260; DB 12; Length 3010;
Best Local Similarity 32.6%; Pred. No. 6.4e-10;
Matches 87; Conservative 29; Mismatches 73; Indels 78; Gaps 10;

QY 23 RPAVIPDEVLVYQFDEME-ASHLPYIEQGMQLAEQFKOKALAF---ASRQNHVSPHY 78
DB 1691 RPAIIPDEVLVYQFDEMECASHLPYIEQGMQLAEQFKOKALGILQTRTKQ----- 1742
QY 79 VPESDARPAIIPDEVLHREPDMEBAFASR-GHVSFAHYVP-----ES 122
DB 1743 -----AEAAVP---VVESKWQALEAFWAKMMNFISGIQYLAGLSTLPGNLAIASLMAF 1793
QY 123 DASQAPYIEQAOVI-----AHQFKVLAFAFRGNHDSPTHVVPESD 165
DB 1794 TASIPTLTHTLTLFNLGCVWVAQIAPPSASAFAFGAGIAGAAVGSIGLKVLDILA 1853
QY 166 AKPAIIPDEVLVYREFD-EMESQHL-----PYIEQGMQL-----AE 201
DB 1854 GYGAGVAGALVAFKMGSEMTEDLVNLPAISPGALVGVCAALIRRHVGPEGAV 1913
QY 202 QFKOKALAFASRGHNVAPTHVTESDA 228
DB 1914 QNMNRLIAFASRGHNVSPHYVPESDA 1940

RESULT 11
QY 0930367 PRELIMINARY; PRT; 3010 AA.
AC 0930367;
I 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MD25;
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;
RT "Characteristics of hepatitis C viral genome associated with disease
RT progression.";
RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
CC -I- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA (BY SIMILARITY).
DB EMBL: AF207766; AAF65956.1; -
DR HSP; P26663; IXP.
DR InterPro: IPR000345; CytC_heme_bind.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR000745; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.

```

DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid_1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR PROSITE: PS00190; CYTOCHROME_C; 1.
 DR PROSITE: PS50507; RDRP_POSITIVE; 1.
 DR PROSITE: PS50521; RDRP_VIRAL; 1.
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 K_m RNA-directed RNA polymerase; Transferrase; Transmembrane.
 K_m RNA-directed RNA polymerase; Transferrase; Transmembrane.
 SQ SEQUENCE 3010 AA; 326507 MW; 9EBAC8ADA4B41DF CRC64;

Query Match 14.3%; Score 260; DB 12; Length 3010;
 Best Local Similarity 32.1%; Pred. No. 6.4e-10;
 Matches 86; Conservative 27; Mismatches 75; Indels 80; Gaps 10;
 QY 23 RPAVIPDREVLVYOEFDMEB-ASHLPYIEQGMLAEQFKOKALAF---ASRQNHVSPTHY 78
 DB 1691 RPVVDPREVLYREFDEMECASHLPYIEQGMLAEQFKOKALGLQTKQAQEA----- 1745
 QY 79 VPESDARPAIIPREVLRHREFDEMEAFASR--GNHVSAPHYVPESDAQAPYIQ----- 133
 DB 1746 ----AAPVESKRAL-----EAFWAKHMMNTISGIQYLAGISTLPGNPAIASLMA 1792
 QY 134--AQVIAHOFKEKVLAF-----ASRGNHDSPTHYVPESDAKPAI----- 170
 DB 1793 FTASITSPLTTOGTLFNLIGWVAQAQAPSAASAFVAGAGIAGAVSGIGKVLVDIL 1852
 QY 171 -----IPREVLYREFDEMEBSOHL--PYIEQGMML-----A 200
 DB 1853 AGYAGVAGALVAFKMTSGEMPSTEDLVNLPAILSPGALVGVCAAILRRHVGPGEA 1912
 QY 201 EOKKALAFASRGNHVAPTHYTESDA 228
 DB 1913 VQMMNRLLIAFASRGNHVSPTHYVPESDA 1940

RESULT 12
 Q9DTES PRELIMINARY; PRT; 3010 AA.
 AC Q9DTES; 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Genome polyploid.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepatitis C virus.
 OC NCBI_TaxID=11103;
 RN [1] SEQUENCE FROM N.A.
 RP STRAIN=HCV145;
 RC Takahashi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,
 RA Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,
 RA Mishiro S.;
 RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients
 with hepatocellular carcinoma: the 'progenitor score' revisited";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 CC - SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND RNA (BY SIMILARITY).
 DR EMBL: AB049092; BAB18805.1; -
 DR HSP: P27958; 1HEF.
 DR InterPro: IPR000345; Cyt_c_heme_bind.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002522; HCV_capsid.
 DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002511; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR00745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR PROSITE: PS00190; CYTOCHROME_C; 2.
 DR PROSITE: PS50507; RDRP_POSITIVE; 1.
 DR PROSITE: PS50521; RDRP_VIRAL; 1.
 DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
 K_m Hydroxylase; Nonstructural protein; Polypeptide;
 K_m RNA-directed RNA polymerase; Transferrase; Transmembrane.
 SQ SEQUENCE 3010 AA; 326586 MW; 573C0F3C55B3F3F4 CRC64;

Query Match 14.3%; Score 260; DB 12; Length 3010;
 Best Local Similarity 31.8%; Pred. No. 6.4e-10;
 Matches 85; Conservative 29; Mismatches 75; Indels 78; Gaps 10;
 QY 23 RPAVIPDREVLVYOEFDMEB-ASHLPYIEQGMLAEQFKOKALAF---ASRQNHVSPTHY 78
 DB 1691 RPVVDPREVLYREFDEMECASHLPYIEQGMLAEQFKOKALGLQTKQAQEA----- 1745
 QY 79 VPESDARPAIIPREVLRHREFDEMEAFASR--GNHVSAPHYVPESDAQAPYIQ----- 133
 DB 1746 ----AAP-----VSESKRALGFWAKHMMNTISGIQYLAGISTLPGNPAIASLMA 1793
 QY 134--AQVIAHOFKEKVLAF-----ASRGNHDSPTHYVPESDAKPAI----- 170
 DB 1794 TASITSPLTTOGTLFNLIGWVAQAQAPSAASAFVAGAGIAGAVSGIGKVLVDILA 1853
 QY 171 -----IDREVLYREFDEMEBSOHL--PYIEQGMML-----AE 201
 DB 1854 GYAGVAGALVAFKMTSGEMPSTEDLVNLPAILSPGALVGVCAAILRRHVGPGEAV 1913
 QY 202 OFKOKALAFASRGNHVAPTHYTESDA 228
 DB 1914 QMMNRLLIAFASRGNHVSPTHYVPESDA 1940

RESULT 13
 Q9J3F4 PRELIMINARY; PRT; 3008 AA.
 AC Q9J3F4; 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 CC


```

Db      1746 -----AAP-----VSESKRALFTFWAKHMMNFISGIQYLGLSTLPGNPAT----- 1787
QY      138 AHQFEKYLAFASRGNNHSPHYVPESDAKPAIIDREVLYREFDEMEESCHLPYIEGGM 197
Db      1788 -----ASMAFTASTISPLTQYT-----LLFNILIGWVAQAQLAPPASASA 1828
QY      198 MLAEOFKOKALAFASRGNNHVAPTHVTESDAKPALVPKEXVLYQOYDEME-----AFA 251
Db      1829 FVGAGIAGAAVGSIGLKVLDIILAGYAGVAGALVAFK-VMSGDMSTEDLVNLLPATIL 1887
QY      252 SRG-----NHVAPTHVYESDASASLPYMDETPAIAGQFEKYLAFASRGNNH 298
Db      1888 SPGALVGVCAALIRRHVGPGEAV-----QMMNRLIAFASRGNNH 1928
QY      299 VSPRHYVPESEPQVVT 315
Db      1929 VSPTHVPESDAARVT 1945

```

JUL 15

```

:K3
:Q90IX3 PRELIMINARY; PRT; 3010 AA.

```

```

AC      Q90IX3;
DT      01-MAY-2000 (Tremblrel. 13, Created)
DT      01-MAY-2000 (Tremblrel. 13, last sequence update)
DT      01-MAR-2003 (Tremblrel. 23, last annotation update)
DE      Genome polyprotein.
OS      Hepatitis C virus.
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC      Hepacivirus.
OX      NCBI_Taxid=11103;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      STRAIN=MD9-2;
RX      MEDLINE=20013325; PubMed=10544098;
RA      Nagayama K., Kurosaki M., Enomoto N., Maekawa S.Y., Miyasaka Y.,
RA      Tazawa J.I., Izumi N., Marumo F., Sato C.;
RT      "time-related changes in full-length hepatitis C virus and hepatitis
RT      activity.";
RL      Virology 263:244-253(1999).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=MD9-2;
RA      Nagayama K., Kurosaki M., Enomoto N., Maekawa S., Miyasaka Y.,
RA      Sakamoto N., Fukuma T., Tazawa J., Izumi N., Marumo F., Sato C.;
RL      Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC      -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
        LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
        PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
        PROTEIN C AND MRNA (BY SIMILARITY).

```

```

DR      EMBL; AF165062; AAD56197.1; -.
DR      HSP; P26663; INS3.
DR      InterPro; IPR001410; DEAD.
DR      InterPro; IPR002522; HCV_capsid.
DR      InterPro; IPR002521; HCV_core.
DR      InterPro; IPR002519; HCV_env.
DR      InterPro; IPR002531; HCV_NS1.
DR      InterPro; IPR002518; HCV_NS2.
DR      InterPro; IPR004109; HCV_NS3.
DR      InterPro; IPR000745; HCV_NS4a.
DR      InterPro; IPR001490; HCV_NS4b.
DR      InterPro; IPR002868; HCV_NS5a.
DR      InterPro; IPR002166; HCV_NS5b.
DR      InterPro; IPR007095; RNA_pol_DS_PS.
DR      InterPro; IPR007094; RNA_pol_PSVir.
DR      Pfam; PF01543; HCV_capsid; 1.
DR      Pfam; PF01542; HCV_core; 1.
DR      Pfam; PF01539; HCV_env; 1.
DR      Pfam; PF01560; HCV_NS1; 1.
DR      Pfam; PF01538; HCV_NS2; 1.
DR      Pfam; PF02907; HCV_NS3; 1.
DR      Pfam; PF01006; HCV_NS4a; 1.

```

```

DR      Pfam; PF01001; HCV_NS4b; 1.
DR      Pfam; PF01506; HCV_NS5a; 1.
DR      Pfam; PF00998; Viral_RdRp; 1.
DR      ProDom; PD166062; HCV_NS1; 1.
DR      SMART; SM00487; DEXDC; 1.
DR      PROSITE; PSS0507; RdRp_POSITIVE; 1.
DR      PROSITE; PSS0521; RdRp_VIRAL; 1.
KW      Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW      Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ      SEQUENCE 3010 AA; 327253 MW; 9F1B0B3F536774FA CRC64;

```

```

Query Match      14.2%; Score 259; DB 12; Length 3010;
Best Local Similarity 31.5%; Pred. No. 7,6e-10;
Matches 84; Conservative 31; Mismatches 74; Indels 78; Gaps 10;

```

```

QY      23 RPAVIDPREVLYOEFDEMEE-ASHLPYIEGQMLAEOFKOKALAF---ASRGNNHVSPTHY 78
Db      1691 RPAVIDPREVLYREFDEMEESCHLPYIEGQMLAEOFKOKALGLLOTATKQ----- 1742
QY      79 VPESDARPAIIDREVLYHREFDEMEEFASR-GNHVSPHYVPESDASQAAPYIEQ---- 133
Db      1743 -----AEAAPV---VSESKRALFTFWAKHMMNFISGIQYLGLSTLPGNPATIASLMAF 1793
QY      134 -AQVIAHQFEKYLAF-----ASRGNNHSTHYVPESDAKPAI----- 170
Db      1794 TASITSPLTQTNTLLENILIGWVAQAQLAPPASASAFVAGIAGAAVGSIGLKVLDIILA 1853
QY      171 -----IPREVLYREFDEMEESCHL--PYIEGQML-----AE 201
Db      1854 GYGAGVAGALVAFKINSGEMSTEDLVNLLPATILSPGALVGVCAALIRRHVGPGEAV 1913
QY      202 QFKOKALAFASRGNNHVAPTHVYESDA 228
Db      1914 QMMNRLIAFASRGNNHVSPTHVPESDA 1940

```

Search completed: July 25, 2003, 13:11:59.
Job time : 100 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: July 25, 2003, 12:55:06 ; Search time 23 Seconds

(without alignments)
719.713 Million cell updates/sec

Title: US-09-491-146a-52
Perfect score: 1818
Sequence: 1 AAHPIYLEQGHLEQFKQK.....KALIEBQGMWMLKSIQ 352

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

1 number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	260	14.3	3010	1	P26662 h genome po
2	258	14.2	3010	1	P26663 h genome po
3	256	14.1	3010	1	P26664 h genome po
4	254	14.1	3010	1	P26665 h genome po
5	254	14.0	3011	1	P26666 h genome po
6	247	13.6	3011	1	P26667 h genome po
7	203.5	11.2	3033	1	P26668 h genome po
8	201.5	11.1	3033	1	P26669 h genome po
9	121.5	6.7	911	1	P26670 h genome po
10	120.5	6.6	5327	1	P26671 h genome po
11	114.5	6.3	912	1	P26672 h genome po
12	114.5	6.3	4473	1	P26673 h genome po
13	113.5	6.2	911	1	P26674 h genome po
14	113.5	6.2	4684	1	P26675 h genome po
15	109.5	6.0	4687	1	P26676 h genome po
16	107.5	5.9	1733	1	P26677 h genome po
17	107	5.9	910	1	P26678 h genome po
18	106.5	5.9	1736	1	P26679 h genome po
19	106	5.8	2748	1	P26680 h genome po
20	104	5.7	964	1	P26681 h genome po
21	103	5.7	3259	1	P26682 h genome po
22	103	5.7	5430	1	P26683 h genome po
23	103	5.7	5938	1	P26684 h genome po
24	101.5	5.6	788	1	P26685 h genome po
25	101.5	5.6	2517	1	P26686 h genome po
26	101	5.6	478	1	P26687 h genome po
27	101	5.6	1036	1	P26688 h genome po
28	101	5.6	3678	1	P26689 h genome po
29	100	5.5	478	1	P26690 h genome po
30	100	5.5	649	1	P26691 h genome po
31	100	5.5	1036	1	P26692 h genome po
32	100	5.5	1036	1	P26693 h genome po
33	100	5.5	1745	1	P26694 h genome po

34	99	5.4	631	1	SYR_HA1N1	Q9huz2 halobacteri
35	98.5	5.4	904	1	AAQ4_CHICK	Q90734 gallus gall
36	98.5	5.4	932	1	IF3A_SCHPO	O74760 schizosach
37	97.5	5.4	468	1	SYE_THERM	P27000 thermus the
38	97.5	5.4	1026	1	YH2O_CAEEL	Q23223 caenorhabdi
39	97	5.3	632	1	CCB2_RABIT	P54288 oryctolagus
40	97	5.3	660	1	CCB2_HUMAN	O08289 homo sapien
41	97	5.3	1756	1	PEPL_HUMAN	O60437 homo sapien
42	96.5	5.3	431	1	RAPH_BACSU	P40771 bacillus su
43	96.5	5.3	562	1	ILVD_STHAM	Q99819 staphylococ
44	96	5.3	2805	1	MAPA_HUMAN	P78559 homo sapien
45	96	5.3	5120	1	PCLO_CHICK	Q9p36 gallus gall

ALIGNMENTS

RESULT 1
ID POLG_HCVUA STANDARD; PRT; 3010 AA.
AC P26662;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polypeptide [Contains: Capsid protein C (core protein) (P22);
DE Envelope glycoprotein E1 (GP12) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7, Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepadivirin)
DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate Japanese) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_Taxid=11116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9108550; PubMed=2175903;
RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,
RA Sugimura T., Shimotohno K.;
RT "Molecular cloning of the human hepatitis C virus genome from
RT Japanese patients with non-A, non-B hepatitis.";
RN Proc. Natl. Acad. Sci. U.S.A. 87:9524-9526(1990).
[2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=9192160; PubMed=1849488;
RA Kato N., Hijikata M., Nakagawa M., Ootsuyama Y., Muraiso K.,
RA Ohkoshi S., Shimotohno K.;
RT "Molecular structure of the Japanese hepatitis C viral genome.";
RL FEBS Lett. 280:325-326(1991).
CC -I- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -I- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polypeptide, commonly with Asp or Glu in the P6
CC position. Cys or Thr in P1 and Ser or Ala in P1'.
CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate +
CC (RNA) (N).
CC -I- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D90208; BAA14233.1; -.

Matches	86;	Conservative	28;	Mismatches	75;	Indels	78;	Gaps	10;
---------	-----	--------------	-----	------------	-----	--------	-----	------	-----

```

OY      23 RPAVLPDEVLVYGFEDMEE-ASHLPYIEGMOGLAEQFKALAF---ASRQNHVSPTHY 78
Db      1691 RPAVLPDEVLVYGFEDMEECASHLPYIEGMOGLAEQFKALGILGTATKOAE- 1745
OY      79 VPEDDARPAIIPDEVILHREDEMEEAASR-GNHVSRANHVPESDASQAPYIEQ---- 133
Db      1746 -----AAP-----VVESKRALEVFQAKMMNFISGIQYLAGLSTLPGRPAISLMAF 1793
OY      134 -AOVIAQFKEKVLAF-----ASRQNHDSPTHYHVPESDAKAPV----- 170
Db      1794 TASITSPYLTONTLTFNLIGWVAALQALRPSAASAFQAGAGLGAAGVSGIGKVLVDILA 1855
OY      171 -----IPDEVLYREFDEMEESQHL--PYIEGMMT-----AE 201
Db      1854 GYGAGVGAGALVAFKVMGSEMPSTEDLVNLLRPAISPGALVGVCAAILRHNVGEGAV 1913
OY      202 QFKQALAFASRGNHVAPTHVYTSDDA 228
Db      1914 QMMNRLIAFASRGNHVSPTHVYPSDDA 1940

RESULT 2
POLG_HCVTW STANDARD; PRT; 3010 AA.
AC P29846;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polypeptide (Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP33) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.7); Protease/helicase NS3 (P70) (Hepadactivin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)}.
OS Hepatitis C virus (isolate Taiwan) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31645;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=92230206; Pubmed=331449;
RA Chen P.-J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
RT "The Taiwanese hepatitis C virus genome: sequence determination and
RT mapping the 5' termini of viral genomic and antigenomic RNA.";
RL Virology 188:102-113(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polypeptide, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEIN:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY 329.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
OR EMBL; M84754; -; NOT_ANNOTATED_CDS.
OR PIR; A40244; GNNVTW.
OR PDB; 1N64; 25-FEB-03.
OR

```

	Query March	14.2%	Score 258;	DB 1;	Length 3010;
	Best Local Similarity	31.1%	Pred. No. 5.5e-09;		
	Matches 85;	Conservative 29;	Mismatches 69;	Indels 90;	Gaps 11;
Dy	23 RPAVPPREVLYQGFDEMEE-AASHLPYTEOGMQLAEQFKOKALAF-----ASRKNHVSPTHY 78	:	:	:	:
Dd	1691 KPAVPBPREVLYQGFDEMEBECASHLPRIEQGMQLAEEQFKKALLLQTATPQAAA----- 1745	:	:	:	:
Oy	79 VPESDARPAIIPREVLHREFDEMEEAFAF-RGNHVSAPHYVPSDSQAAPYTEQ--- 133	:	:	:	:
Dd	1746 -----AAP-----VVESKRTELEAFAMNMNMFISIGIQYLACLSTLPGRPAIASIMAF 1793	:	:	:	:
Oy	134 -AQVIHQFEXKALFASRGNHDPPTYHVPSDAKP-----I 170	:	:	:	:
Dd	1794 TASTSPLTSTTOSTLLFNILG-----WVAAGLPRPAASAFLVGAGIAGAAVSIGLGKV 1847	:	:	:	:
Oy	171 IPDR-----EVLYREFDEMSESOHL--PYIEOQMNL----- 199	:	:	:	:
Dd	1848 LVDNVAGYGAGVALVAFKMVMSEMPSTEDLVNLLRALSLPSGLVGVCAILLRHND 1907	:	:	:	:
Oy	200 ---AEQFKOKALAFASRGNHVADPTHYTESDA 228	:	:	:	:
Dd	1908 PGEAGVOMNMRLIAFASRGNHVSPTHYVESDA 1940	:	:	:	:
	RESULT 3				
POLG_HCVBK	POLG_HCVBK STANDARD; PRT; 3010 AA.				
ID	P26663;				
AC	01-AUG-1992 (Rel. 23, Created)				
DT	01-AUG-1992 (Rel. 23, Last sequence update)				
DT	15-SEP-2003 (Rel. 42, Last annotation update)				
DE	Genome polypeptide [Contains: Capsid protein C (core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitisin) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P55); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].				
DE	Hepatitis C virus [isolate BK] (HCV).				
OS	Virusess; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; OC Hepacivirus.				
OC	NCB1_TaxID=11105;				
CX	[1]				
RN	SEQUENCE FROM N.A.				
RP	MEDLINE=91140696; PubMed=1847440;				
RA	Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J., Onishi E., Andoh T., Yoshida I., Okayama H.;				
RA	"Structure and organization of the hepatitis C virus genome isolated from human carriers.";				
RT	J. Virol. 65:1105-1113(1991).				
RL	[2]				
RN	SEQUENCE OF 1487-1500.				
RP	MEDLINE=96235224; PubMed=8647104;				
RA	Borowski P., Helland M., Oehlmann K., Becker B., Kornetky L.;				
RT	"Non-structural proteiin 3 of hepatitis C virus inhibits phosphorylation mediated by CAMP-dependent protein kinase.";				
RL	Eur. J. Biochem. 237:611-618(1996).				
RN	[3]				
RN	X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.				
RP	MEDLINE=97015088; PubMed=8861916;				
RA	Lowe R.A., Perge H.E., Wickersham J.A., Hostomsky Z., Hebluka N., Wooman E.W., Adachi T., Hostomska Z.;				
RA	"The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site.";				
RT	Cell 87:331-342(1996).				
RL	[4]				
RP	X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.				
RA	MEDLINE=96227846; PubMed=8568891;				
RA	Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M., Steinhuebler C., Tomei L., de Francesco R., Xuo L.C., Chen Z.;				
RT	"Complex of NS3 protease and NS4A peptide of BK strain hepatitis C"				

RL Protein Sci. 7:837-847(1998).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate +
CC (RNA)(N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: MS8135; AAA72945.1; -
DR PIR: A38465; GNVVTC.
DR PDB: 1A10; 25-MAR-98.
DR PDB: 1JXP; 14-JAN-98.
DR PDB: 1NS3; 08-APR-98.
DR PDB: 1C2P; 15-NOV-00.
DR PDB: 1CSJ; 08-NOV-99.
DR PDB: 1GX5; 09-APR-02.
DR PDB: 1GX6; 10-APR-02.
DR PDB: 1QUV; 26-JUN-00.
DR PDB: 8OHM; 20-APR-99.
DR MEROPS: S29.001; -
DR -----
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002165; HCV_RdRp.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR Pfam: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
REMOVED FROM CAPSID PROTEIN C BY THE
CELLULAR AMINOPEPTIDASE.
CAPSID PROTEIN C (POTENTIAL).
MATRIX PROTEIN (POTENTIAL).
MAJOR ENVELOPE PROTEIN E (POTENTIAL).
NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
PROTEASE/HELICASE NS3 (POTENTIAL).

FT CHAIN 1616 1662
FT CHAIN 1663 2013
FT CHAIN 2014 3010
FT TRANSMEM 347 369
FT ACT_SITE 1083 1083
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 136 196
FT CARBOHYD 209 209
FT CARBOHYD 224 224
FT CARBOHYD 250 250
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 532 532
FT CARBOHYD 540 540
FT CARBOHYD 556 556
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
FT CARBOHYD 2041 2041
FT CARBOHYD 2077 2077
FT CARBOHYD 2240 2240
FT CARBOHYD 2529 2529
FT CARBOHYD 2788 2788
FT STRAND 1031 1035
FT HELIX 1039 1047
FT STRAND 1050 1050
FT STRAND 1059 1063
FT STRAND 1068 1074
FT TURN 1075 1076
FT STRAND 1077 1081
FT HELIX 1082 1085
FT TURN 1086 1087
FT STRAND 1090 1092
FT TURN 1093 1094
FT STRAND 1095 1097
FT STRAND 1101 1103
FT TURN 1104 1107
FT STRAND 1108 1112
FT STRAND 1120 1120
FT STRAND 1122 1122
FT STRAND 1129 1133
FT TURN 1135 1136
FT STRAND 1139 1144
FT STRAND 1149 1157
FT HELIX 1158 1161
FT TURN 1162 1163
FT TURN 1165 1166
FT STRAND 1168 1171
FT TURN 1172 1174
FT STRAND 1175 1186
FT TURN 1187 1188
FT STRAND 1189 1197
FT HELIX 1198 1202
FT TURN 1203 1204
FT STRAND 1680 1688
SQ SEQUENCE 3010 AA; 327189 MW; F8422D5ECFDFD9C CRC64;
Query Match 14.1%; Score 256; DB 1; Length 3010;
Best Local Similarity 31.1%; Pred. No. 7,4e-09;
Matches 83; Conservative 31; Mismatches 75; Indels 78; Gaps 10;
QY 23 RPATVDPREVLVYQFDEMEK-ASHLPYIEQGMQALAEQFKQALAF---ASRQNHVSPTHY 78
DB 1691 RPAIVPDELLVYQFDEMEKASHLPYIEQGMQALAEQFKQALGLQATKQAA----- 1745
QY 79 VPESDAPRALIPDEVLHREFDEMEKAFASR-GHVSPPAHVVPSSDSQAAPYIEQ----- 133

Db 1746 -----AAP-----VSESKRALETFWAKHMMNFISGLOYLAGLSTLGNPAIASLMAF 1793
Qy 134 -AQVIAHOFKEXLAF-----ASRGNDSPTHVYPSDAKPAI----- 170
Db 1794 TBSITSPLTQTSLNLNIGWVAQLAPASASAVGVAGIAAAGSGLKVLVDILA 1853
Qy 171 -----IPDREVLRYREDEMEESQHL--PYIEQGMML-----AE 201
Db 1854 GYGAGVAGALVAFKVGSGEMPTSEDLVNLPLPALSFGALVGVCAAILRHVGPGECAV 1913
Qy 202 OFKOKALAFASRGNHVAPTHVYTESDA 228
Db 1914 QMNRLLAFASRGNHVSPTHVYTESDA 1940
RESULT 4
POLG_HCVUT STANDARD: PRT; 3010 AA.
AC 000269;
01-APR-1993 (Rel. 25, Created)
01-APR-1993 (Rel. 25, Last sequence update)
15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polypeptide (contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepaticivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)).
OS Hepatitis C virus (isolate HC-UT) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_Taxid=31642;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=92295714; PubMed=1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RT infected individuals."
RT Virus Res. 23:39-53 (1992).
CC -I- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -I- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polypeptide, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA) (N).
CC -I- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA.
CC -I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: D11168; BAA01943.1; -
CC PIR: A45573; A45573.
CC PDB: 1A1Q; 25-MAR-98.
CC PDB: 1UXP; 14-JAN-98.
CC MEROPS: S29.001; -
CC MEROPS: U39.001; -
CC InterPro: IPR001410; DEAD.
CC InterPro: IPR002542; HCV_capsid.
CC InterPro: IPR002521; HCV_core.

DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4A.
DR InterPro: IPR001490; HCV_NS4B.
DR InterPro: IPR002868; HCV_NS5A.
DR InterPro: IPR002166; HCV_NS5B.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4A; 1.
DR Pfam: PF01001; HCV_NS4B; 1.
DR Pfam: PF01506; HCV_NS5A; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR KEGG: K01001; Transferase; RNA-directed RNA polymerase;
DR KEGG: K01002; Envelope protein; Helicase; ATP-binding;
DR KEGG: K01003; Transmembrane; Nonstructural protein; Hydrolyase; Serine protease;
DR KEGG: K01004; 3D-structure.
KW INT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 729
FT CHAIN 730 1006
FT CHAIN 1007 1615
FT CHAIN 1616 1862
FT CHAIN 1863 2013
FT CHAIN 2014 3010
FT TRANSMEM 347 369
FT ACT_SITE 1083 1083
FT ACT_SITE 1107 1107
FT ACT_SITE 1165 1165
FT NP_BIND 1230 1237
FT SITE 1316 1319
FT CARBOHYD 136 196
FT CARBOHYD 209 209
FT CARBOHYD 234 234
FT CARBOHYD 250 250
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 532 532
FT CARBOHYD 540 540
FT CARBOHYD 556 556
FT CARBOHYD 576 576
FT CARBOHYD 623 623
FT CARBOHYD 645 645
FT CARBOHYD 2041 2041
FT CARBOHYD 2077 2077
FT CARBOHYD 2240 2240
FT CARBOHYD 2529 2529
FT CARBOHYD 2788 2788
SQ SEQUENCE 3010 AA; 326573 MW; 94A1CT7435D642BB CRC64;
Query Match 14.1%; Score 256; DB 1; Length 3010;
Best Local Similarity 31.5%; Pred. No. 7,4e-09;
Matches 84; Conservative 31; Mismatches 74; Indels 78; Gaps 10;
Qy 23 RPAVIPPREVLYQFDEMEE-ASHLPYIEQGMQALAEQFKOKALF---ASRQNHVSPTHY 78
Db 1691 RPAVPPREVLYQFDEMEECASHLPYIEQGMQALAEQFKOKALGLLOTATKQ----- 1742

QY 23 RPAVDPREVLYOEFDMEEAS-HPYIEQGMOLAEQFKOKALAF---ASRQNHVSPTHY 78
 Db 1691 KPAIPREVLYREFDEMEESCHLPYIEQGMMLAEQFKOKALGLQOTASRQAEV----- 1745
 QY 79 VPESDARPALIPREVLYREFDEMEEAFASR-GNHVSPPHYVESDASQAAPYIEQ---- 133
 Db 1746 ----TAPAV-----QTNWQKLETFWAGHMNFISQIYLAGLSTLPQNPALIASLMAF 1793
 QY 134 -AOVIAHQFKEKVLAF-----ASRGNHDSPTHYVPESDAPAI----- 170
 Db 1794 TAAVTSPLTTSQTLFPIILGQWVAQAOLAPGATATFVGAAGALIGVGLKVLIDILA 1853
 QY 171 -----IPDREVLYREFDEMEESQHT--PYIEQGMML-----AE 201
 Db 1854 GYGAGVAGALVARKINSQGEVPTEDLVNLLPALSPGALVGVVCAILRRHYGPBGAV 1913
 QY 202 QFKOKALAFASRGNHVAPTHYVESDA 228
 Db 1914 QNMNRLIAFASRGNHVSPTHYVESDA 1940

RESULT 6
 POLG_HCVH STANDARD; PRT; 3011 AA.
 ID POLG_HCVH P27958;
 AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein (Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepatitisin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 NC NCBI_TaxID=11108;
 RN (1)
 RN SEQUENCE FROM N.A.
 RA MEDLINE=92052256; Pubmed=1658800;
 RX Inchauspe G., Zebadee S., Lee D.H.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RT "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 [2]
 X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 h. MEDLINE=97331322; Pubmed=9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RT "Structure of the hepatitis C virus RNA helicase domain";
 RL Nat. Struct. Biol. 4:463-467(1997).
 [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE=98154321; Pubmed=9493270;
 RA Kim J.L., Morgensstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.;
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding";
 RL Structure 6:89-100(1998).
 CC -1- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -1- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -1- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -1- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVITY ROLE.
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.

CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA} (N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND RNA.
 CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
 CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U29.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M67463; AAA45534.1; -;
 CC PIR; A36814; GNMVCH.
 CC PDB; 1HE1; 25-NOV-98.
 CC PDB; 1AIV; 16-FEB-99.
 CC PDB; 1AIR; 17-JUN-98.
 CC MEROPS; S29.001; -;
 CC MEROPS; S29.001; -;
 CC TRANSFAC; T04155; -;
 CC InterPro; IPR001410; DEAD.
 CC InterPro; IPR002512; HCV_capsid.
 CC InterPro; IPR002521; HCV_core.
 CC InterPro; IPR002519; HCV_core.
 CC InterPro; IPR002531; HCV_NS1.
 CC InterPro; IPR002518; HCV_NS2.
 CC InterPro; IPR004109; HCV_NS3.
 CC InterPro; IPR000745; HCV_NS4A.
 CC InterPro; IPR001490; HCV_NS4B.
 CC InterPro; IPR002868; HCV_NS5A.
 CC InterPro; IPR002166; HCV_RdRP.
 CC InterPro; IPR001650; Helicase_C.
 CC InterPro; IPR007095; RNA_pol_DS_PS.
 CC InterPro; IPR007094; RNA_pol_PSVlt.
 CC Pfam; PF01543; HCV_capsid; 1.
 CC Pfam; PF01542; HCV_core; 1.
 CC Pfam; PF01539; HCV_env; 1.
 CC Pfam; PF01560; HCV_NS1; 1.
 CC Pfam; PF01538; HCV_NS2; 1.
 CC Pfam; PF02907; HCV_NS3; 1.
 CC Pfam; PF01006; HCV_NS4a; 1.
 CC Pfam; PF01001; HCV_NS4b; 1.
 CC Pfam; PF01506; HCV_NS5a; 1.
 CC Pfam; PF00271; helicase_C; 1.
 CC Pfam; PF00998; Viral_RdRP; 1.
 CC ProDom; PD186062; HCV_NS1; 1.
 CC SMART; SM00487; DEXDC; 1.
 CC PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 CC Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 CC Transmembrane; Nonstructural protein; Hydrolyase; Serine protease;
 CC 3D-structure.
 CC INIT_MER 1 1
 CC CHAIN 1 191
 CC CHAIN 192 383
 CC CHAIN 384 746
 CC CHAIN 747 809
 CC CHAIN 810 1026
 CC CHAIN 1027 1657
 CC CHAIN 1658 1711
 CC CHAIN 1712 1972
 CC CHAIN 1973 2420
 CC CHAIN 2421 3011
 CC TRANSMEM 347 369
 CC ACT_SITE 1083 1083
 CC ACT_SITE 1107 1107
 CC ACT_SITE 1165 1165
 CC REMOVED FROM CAPSID PROTEIN C BY THE
 CC CELLULAR AMINOPEPTIDASE.
 CC CAPSID PROTEIN C.
 CC ENVELOPE GLYCOPROTEIN E1.
 CC ENVELOPE GLYCOPROTEIN E2.
 CC PROTEIN P7.
 CC NONSTRUCTURAL PROTEIN NS2.
 CC PROTEASE/HELICASE NS3.
 CC NONSTRUCTURAL PROTEIN NS4A.
 CC NONSTRUCTURAL PROTEIN NS4B.
 CC NONSTRUCTURAL PROTEIN NS5A.
 CC NONSTRUCTURAL PROTEIN NS5B.
 CC POTENTIAL.
 CC CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC CHARGE RELAY SYSTEM (BY SIMILARITY).
 CC CHARGE RELAY SYSTEM (BY SIMILARITY).

FT	NP BIND	1230	1237	ATP (POTENTIAL).
FT	SITE	1316	1319	DECH BOX.
FT	CARBOHYD	196	196	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	209	209	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	234	234	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	305	305	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	417	417	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	423	423	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	430	430	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	448	448	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	476	476	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	532	532	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	540	540	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	556	556	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	576	576	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	623	623	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	645	645	N-LINKED (GLCNAC. . .)
FT	STRAND	1224	1226	N-LINKED (GLCNAC. . .)
FT	TURN	1232	1233	N-LINKED (GLCNAC. . .)
FT	TURN	1236	1238	N-LINKED (GLCNAC. . .)
FT	HELIX	1239	1246	N-LINKED (GLCNAC. . .)
FT	TURN	1247	1248	N-LINKED (GLCNAC. . .)
FT	HELIX	1251	1255	N-LINKED (GLCNAC. . .)
FT	TURN	1258	1271	N-LINKED (GLCNAC. . .)
FT	TURN	1272	1272	N-LINKED (GLCNAC. . .)
FT	STRAND	1277	1280	N-LINKED (GLCNAC. . .)
FT	TURN	1281	1282	N-LINKED (GLCNAC. . .)
FT	STRAND	1283	1285	N-LINKED (GLCNAC. . .)
FT	STRAND	1291	1295	N-LINKED (GLCNAC. . .)
FT	HELIX	1296	1301	N-LINKED (GLCNAC. . .)
FT	TURN	1302	1303	N-LINKED (GLCNAC. . .)
FT	STRAND	1312	1316	N-LINKED (GLCNAC. . .)
FT	TURN	1317	1319	N-LINKED (GLCNAC. . .)
FT	HELIX	1323	1335	N-LINKED (GLCNAC. . .)
FT	TURN	1336	1340	N-LINKED (GLCNAC. . .)
FT	STRAND	1343	1347	N-LINKED (GLCNAC. . .)
FT	TURN	1352	1353	N-LINKED (GLCNAC. . .)
FT	STRAND	1360	1361	N-LINKED (GLCNAC. . .)
FT	STRAND	1362	1366	N-LINKED (GLCNAC. . .)
FT	STRAND	1368	1368	N-LINKED (GLCNAC. . .)
FT	STRAND	1373	1375	N-LINKED (GLCNAC. . .)
FT	TURN	1376	1377	N-LINKED (GLCNAC. . .)
FT	STRAND	1378	1380	N-LINKED (GLCNAC. . .)
FT	HELIX	1382	1385	N-LINKED (GLCNAC. . .)
FT	STRAND	1389	1393	N-LINKED (GLCNAC. . .)
FT	HELIX	1397	1409	N-LINKED (GLCNAC. . .)
FT	TURN	1410	1411	N-LINKED (GLCNAC. . .)
FT	STRAND	1414	1417	N-LINKED (GLCNAC. . .)
FT	TURN	1419	1420	N-LINKED (GLCNAC. . .)
FT	STRAND	1432	1436	N-LINKED (GLCNAC. . .)
FT	TURN	1438	1439	N-LINKED (GLCNAC. . .)
FT	STRAND	1450	1453	N-LINKED (GLCNAC. . .)
FT	STRAND	1456	1463	N-LINKED (GLCNAC. . .)
FT	STRAND	1471	1478	N-LINKED (GLCNAC. . .)
FT	STRAND	1480	1480	N-LINKED (GLCNAC. . .)
FT	HELIX	1481	1488	N-LINKED (GLCNAC. . .)
FT	TURN	1489	1490	N-LINKED (GLCNAC. . .)
FT	STRAND	1497	1501	N-LINKED (GLCNAC. . .)
FT	STRAND	1507	1507	N-LINKED (GLCNAC. . .)
FT	STRAND	1511	1511	N-LINKED (GLCNAC. . .)
FT	HELIX	1514	1527	N-LINKED (GLCNAC. . .)
FT	HELIX	1532	1544	N-LINKED (GLCNAC. . .)
FT	STRAND	1550	1550	N-LINKED (GLCNAC. . .)
FT	HELIX	1555	1564	N-LINKED (GLCNAC. . .)
FT	HELIX	1570	1578	N-LINKED (GLCNAC. . .)
FT	TURN	1579	1580	N-LINKED (GLCNAC. . .)
FT	HELIX	1584	1597	N-LINKED (GLCNAC. . .)
FT	TURN	1598	1598	N-LINKED (GLCNAC. . .)
FT	HELIX	1606	1611	N-LINKED (GLCNAC. . .)
FT	TURN	1614	1618	N-LINKED (GLCNAC. . .)
FT	STRAND	1622	1623	N-LINKED (GLCNAC. . .)
FT	STRAND	1627	1627	N-LINKED (GLCNAC. . .)
FT	STRAND	1635	1636	N-LINKED (GLCNAC. . .)

FT HELIX 1640 1652
SO SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;
Query Match 13.6%; Score 247; DB 1; Length 3011;
Best Local Similarity 32.2%; Pred. No. 28-08;
Matches 86; Conservative 26; Mismatches 77; Indels 78; Gaps 11;

QY 23 RPAVIDPREVLVYDFEDEMEEAS-HLPYIEGQMLAEQFKOKALAF---ASRONHVSPTHY 78
1691 KPAIIPREVLVYDFEDEMEECSOHLPIEQQMLAEQFKOKALGLQTAASHAEV----- 1745
Db 79 VPESDAPRAIIPREVLVHREFDEMEEPFASR-GNHVSPAHVYESDASQAAPYIEQ---- 133
1746 -----ITPAV-----QTNQKLEVFQWAGHMMNFISGIOYLAGLSTLPGNPAIASIMAF 1793
QY 134 -AQVIAHOFKEXVIAF-----ASRGNDSPTHVVPESDAKPAI----- 170
1794 TAAVTSPLTTGGQTLLENILGWAALAPGAATATVAGAGLAAALDSVGLKAVLDILA 1853
Db 171 -----IPREVLYREFDEMEESSQH-LPYI-----EQGMMLAE----- 201
1854 GYGAGVAGALVAFKINSGEVPTEDLVNLLPALSGLALVGVVFAISLRRYVPEGAV 1913
QY 202 QFKOKALAPASRGNHVAPHYVYESDA 228
Db 1914 QNMNRLIAFASRGNHVSPHYVYESDA 1940

RESULT 7
POLG_HCVJ6
ID POLG_HCVJ6 STANDARD; PRT; 3033 AA.
AC P26660;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein (containing: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protease/helicase NS3 (P70) (Hepatitisin)
DE (EC 3.4.22.-); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)).
OS Hepatitis C virus (isolate HC-06) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_Taxid=11113;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92044440; PubMed=1658196;
RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Litzuka H.,
RA Machida A., Miyakawa Y., Mayumi M.;
RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
RT from a human carrier: comparison with reported isolates for conserved
RT and divergent regions.";
RL J. Gen. Virol. 72:2697-2704(1991).
CC -I- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -I- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the p6
CC position, Cys or Thr in p1 and Ser or Ala in p1'.
CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -I- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA.
CC -I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

FT CARBOHYD 2811 2811 N-LINKED (GLCNAC...) (POTENTIAL)
SQ SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;
Query Match 11.2%; Score 203.5; DB 1; Length 3033;
Best Local Similarity 44.1%; Pred. No.2e-05;
Matches 60; Conservative 10; Mismatches 37; Indels 29; Gaps 5;

Oy 229 KPALVPDKVELVYQOYDEMEEFASRGNHVAPTHYVESDAS---ASLPYNDETRAIAGOF 285
Db 1619 KPTLVGPTPLLYR-----LQSVTNEVLTHTPVTKYATNCQADLEWVTSITWVLVAGV 1670
Oy 286 KEKVLASASGN-----HVSFRAHVPESEPOVYVTPKKEILYEAFFDEMEF-ASGAAL 336
Db 1671 LAAYAAACATGCVCIIGRLHVNR-----AVVAPDKVELVYEAFFDEMEECASRAAL 1721
Oy 337 IEEGQMAEMLKSKIQ 352
Db 1722 IEEGQRIAEMLKSKIQ 1737

RESULT 8
POLG_HCVJ8
ID POLG_HCVJ8 STANDARD; PRT; 3033 AA.
AC P26661;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Envelope glycoprotein [Contains: Capsid protein C (core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepadivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-J8) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_Taxid=11115;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kuzai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
RA Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes.";
RL Virology 186:331-341(1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate + N diphosphate +
CC (RNA) (N).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY 529.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage of and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL, D10988; BAA01761.1; -
DR PIR, A40250; GNMVJ8.
DR HSSP, P27958; 1HE1.
DR MEROPS: S29.001; -.

```

DR MEROPS: U39_001: -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_NS5b.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00998; Viral_RdRp; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDc; 1.
DR PolyProtein: Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
protein; Hydrolyase; Serine protease.
KW Transmembrane; Nonstructural
INIT_MET 1 1
FT CHAIN 1 115
FT CHAIN 116 191
FT CHAIN 192 383
FT CHAIN 384 733
FT CHAIN 734 1010
FT CHAIN 1011 1619
FT CHAIN 1620 1866
FT CHAIN 1867 2017
FT CHAIN 2018 3033
FT TRANSMEM 347 369
FT ACT_SITE 1087 1087
FT ACT_SITE 1111 1111
FT ACT_SITE 1169 1169
FT NP_BIND 1234 1241
FT SITE 1320 1323
FT CARBOHYD 196 196
FT CARBOHYD 209 209
FT CARBOHYD 233 233
FT CARBOHYD 299 299
FT CARBOHYD 305 305
FT CARBOHYD 417 417
FT CARBOHYD 423 423
FT CARBOHYD 430 430
FT CARBOHYD 448 448
FT CARBOHYD 477 477
FT CARBOHYD 534 534
FT CARBOHYD 542 542
FT CARBOHYD 558 558
FT CARBOHYD 578 578
FT CARBOHYD 627 627
FT CARBOHYD 649 649
FT CARBOHYD 1091 1091
FT CARBOHYD 2038 2038
FT CARBOHYD 2359 2359
FT CARBOHYD 2811 2811
SO SEQUENCE 3033 AA; 330177 MW; 1A173E7E381FDDA CRC64;
Query Match. 11.1%; Score 201.5; DB 1; Length 3033;
Best Local Similarity 46.2%; Pred. No. 2.6e-05;
Matches 60; Conservative 13; Mismatches 40; Indels 17; Gaps 5;

Db 1619 KPTLTGPPPLR-----LCAVTNETHPTKTYATCMQADLIMSSWLACGV 1670
QY 286 KEKVLAFASRGNHVS--PRHYVESEPQVVTTPDXEILYEAFDEME-ASKAALIEGQR 342
Db 1671 LAAVAAYCLATGCTCIIRLHND---RVVAPDKEIILEYAFDEMEFCSKALIEGQR 1727
QY 343 MAEMLKSKIQ 352
Db 1728 MAEMLKSKIQ 1737
RESULT 9
AAC4 RAT STANDARD; PRT; 911 AA.
ID AAC4 RAT
AC 09OXQ0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Alpha-actinin 4 (Non-muscle alpha-actinin 4) (F-actin cross linking
protein).
GN ACTN4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20139748; PubMed=10673389;
RA El-Husseini A.E.-D., Kwasnicka D., Yamada T., Hirahashi S.,
Vincent S.R.;
"BERP, a novel ring finger protein, binds to alpha-actinin-4";
Biochem. Biophys. Res. Commun. 267:906-911(2000).
- FUNCTION: F-ACTIN CROSS-LINKING PROTEIN WHICH IS THOUGHT TO ANCHOR
ACTIN TO A VARIETY OF INTRACELLULAR STRUCTURES. THIS IS A BUNDLING
PROTEIN.
CC - SUBUNIT: HOMODIMER, ANTIPARALLEL (BY SIMILARITY). BINDS BERP AT
CC THE N-TERMINUS.
CC - SUBCELLULAR LOCATION: Cytoplasmic.
CC - SIMILARITY: Contains 1 actin-binding domain.
CC - SIMILARITY: Contains 2 calponin-homology (CH) domains.
CC - SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC - SIMILARITY: Contains 4 spectrin repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb.ch/announce/>
or send an email to license@isb.sib.ch).
CC -----
DR EMBL: AF190909; AAF20064.1; -;
DR PIR: JC7186; JC7186.
DR HSP: Q01082; 1BRK.
DR InterPro: IPR001589; Actbind_actnin.
DR InterPro: IPR001715; Calponin-like.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR002017; Spectrin.
DR Pfam: PF00307; CH; 2.
DR Pfam: PF00036; ehand; 2.
DR Pfam: PF00435; spectrin; 4.
DR ProDom: PD000012; EF-hand; 1.
DR SMART: SM00033; CH; 2.
DR SMART: SM00054; EFh; 2.
DR SMART: SM00150; SPEC; 3.
DR PROSITE: PS00019; ACTININ_1; 1.
DR PROSITE: PS00020; ACTININ_2; 1.
DR PROSITE: PS50021; CH; 2.
DR PROSITE: PS00018; EF_HAND; 1.
KW Actin-binding; Calcium-binding; Repeat; Multigene family;
DOMAIN 1 269 ACTIN-BINDING (BY SIMILARITY).
FT

DR PROSITE; PS00018; EF_HAND; 1.
 DR PROSITE; PS00019; ACTININ_1; 1.
 DR PROSITE; PS00020; ACTININ_2; 1.
 DR PROSITE; PS00021; CH; 2.
 KW Actin-binding; Calcium-binding; Repeat; Multigene family.
 FT DOMAIN 1 267 ACTIN-BINDING (BY SIMILARITY).
 FT DOMAIN 51 155 CH 1.
 FT DOMAIN 164 267 CH 2.
 FT DOMAIN 178 193 POLYPHOSPHOINOSITIDE (PIP2)-BINDING (POTENTIAL).
 FT REPEAT 294 404 SPECTRIN 1.
 FT REPEAT 414 519 SPECTRIN 2.
 FT REPEAT 529 640 SPECTRIN 3.
 FT REPEAT 650 753 SPECTRIN 4.
 FT CA_BIND 779 790 EF-HAND 1 (POTENTIAL).
 FT CA_BIND 820 831 EF-HAND 2 (POTENTIAL).
 SO SEQUENCE 912 AA; 104976 MW; 07AA9C92AC228B5A CRC64;

Query Match 6.3%; Score 114.5; DB 1; Length 912;
 at Local Similarity 19.9%; Pred. No. 2.8;
 tches 80; Conservative 65; Mismatches 124; Indels 133; Gaps 20;

QY 7 LEOGHALAFOFKOKA-----LRPAVIPREVLVYQEFDEMEE--ASHL 46
 DB 409 LERLDHLAEKFRKOKASIHAWTDGKEAMLKORDYETATLSDIKALIRKHEAFESDLAAHQ 468
 QY 47 PYIEGOMLAEOFKO-----KALAFASRGNHV-SPTHYVESDARPAIIPREV 94
 DB 469 DVEQJAAIAIQENEDYDSHNVTROCKICDQMDNLSLTH-----SRKDALEKTEK 522
 QY 95 LHRFEDMEAFASRGNHVSPAHYVESDASQAAPYIEQAOVIAHOFKEKVALAFASRGNH 154
 DB 523 QLETTDQHLLEYAKRA-----APFNMMESAMEDLDQMFVHTI-EEIGLSIA--H 571
 QY 155 DSEPTHVPESSD-AKPAIIPREVLYREFDEMEBSOHL-----PY-----I 193
 DB 572 DQFKSTLPADDEREMIL-----AIHKEAQRIAESNHIKLSGSPYTTVTPQIINSKWEKV 627
 QY 194 EQ-----GMMLAEQFKOKA-----LAFASRGNHVAP----- 219
 DB 628 QQLVPRKDHALLEBQSKQSNHLEHRQFASQANMGPMIOTKMEIIGRISIMNGTLEDO 687
 QY 220 -THVYTES-----DAKPALVPDKEVLYQDYDEMEAFASRGNHVAPTHVVE----- 265
 DB 688 LSHLQYERSIVDYKPSL-----DLLEQHQQLIQEALIFPNKH-----TNYIMEHIRGVMEQL 740
 QY 266 -SDASASLPYME--TRAIACQFKEKVALAFASRGNHVSPRH 303
 741 LTTIARTINEVENQILTRDAKGISQEQMGEFRASFHFDKDH 782

RESULT 12
 PLE1_CRIGR STANDARD; PRT; 4473 AA.
 AC 09J155;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Plectin 1 (PLTN) (PCN) (300-kDa intermediate filament-associated
 DE protein) (IFAP300) (Fragment).
 GN PLECT1.
 OS Cricetus griseus (Chinese hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Cricetulus.
 OX NCBI_TaxID=10029;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20334248; PubMed=10873583;
 RA Clubb B.H., Chou Y.-H., Herrmann H., Svickina T.M., Borisy G.G.,
 RA Goldman R.D.;
 RT "The 300-kDa intermediate filament-associated protein (IFAP300) is a
 hamster plectin ortholog.";

RL Biochem. Biophys. Res. Commun. 273:183-187(2000).
 RN {2}
 RP PHOSPHORYLATION.
 RX MEDLINE=96215219; PubMed=8626512;
 RA Malecz N., Foisner R., Stadler C., Wiche G.;
 RT "Identification of plectin as a substrate of p34cdc2 kinase and
 RT mapping of a single phosphorylation site.";
 RL J. Biol. Chem. 271:8203-8208(1996).
 CC - FUNCTION: INTERLINKS INTERMEDIATE FILAMENTS WITH MICROTUBULES AND
 CC MICROFILAMENTS AND ANCHORS INTERMEDIATE FILAMENTS TO DESOMES OR
 CC HEMIDESOMES. MAY BE INVOLVED NOT ONLY IN THE CROSSLINKING AND
 CC STABILIZATION OF CYTOSKELETAL INTERMEDIATE FILAMENTS NETWORK, BUT
 CC ALSO IN THE REGULATION OF THEIR DYNAMICS.
 CC - SUBUNIT: HOMODIMER OR HOMOTETRAMER.
 CC - DOMAIN: THE N-TERMINUS INTERACTS WITH ACTIN. THE C-TERMINUS WITH
 CC VIMENTIN, DESMIN, GAP, CYTOKERATINS, LAMIN B; WHEREAS BOTH THE N-
 CC AND THE C-TERMINUS CAN BIND INTERIN BETA-4.
 CC - PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM
 CC INTERMEDIATE FILAMENTS DURING MITOSIS.
 CC - SIMILARITY: Contains 1 actin-binding domain.
 CC - SIMILARITY: Contains 2 calponin-homology (CH) domains.
 CC - SIMILARITY: Contains 32 plectin repeats.
 CC - SIMILARITY: Contains 4 spectrin repeats.
 CC - SIMILARITY: Belongs to the plakin or cytolinker family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF260753; AAF70372.1; -
 DR HSSP; Q01082; IBRK.
 DR InterPro; IPR001589; Actbind_actin.
 DR InterPro; IPR001715; Calponin-like.
 DR InterPro; IPR001101; Plectin repeat.
 DR InterPro; IPR002017; Spectrin.
 DR Pfam; PF00307; CH; 2.
 DR Pfam; PF00681; Plectin; 20.
 DR SMART; SM00033; CH; 2.
 DR SMART; SM00250; PLEC; 34.
 DR PROSITE; PS00019; ACTININ_1; PARTIAL.
 DR PROSITE; PS00020; ACTININ_2; FALSE_NEG.
 DR PROSITE; PS00021; CH; 2.
 KW Coiled coil; Repeat; Structural protein; Cytoskeleton; Actin-binding;
 KW Phosphorylation.
 FT NON_TER 1 1
 FT DOMAIN <1 1259
 FT DOMAIN 1260 2544 GLOBULAR 1.
 FT DOMAIN 2545 4473 CENTRAL FIBROUS ROD DOMAIN.
 FT DOMAIN <1 192 GLOBULAR 2.
 FT DOMAIN <1 74 ACTIN-BINDING.
 FT DOMAIN 87 189 CH 1.
 FT REPEAT 449 508 SPECTRIN 1.
 FT REPEAT 529 613 SPECTRIN 2.
 FT REPEAT 626 719 SPECTRIN 3.
 FT REPEAT 1104 1204 SPECTRIN 4.
 FT REPEAT 1258 2548 SPECTRIN 5.
 FT REPEAT 2615 2652 COILED COIL (POTENTIAL).
 FT REPEAT 2653 2650 PLECTIN 1.
 FT REPEAT 2652 2650 PLECTIN 2.
 FT REPEAT 2691 2728 PLECTIN 3.
 FT REPEAT 2729 2766 PLECTIN 4.
 FT REPEAT 2770 2804 PLECTIN 5.
 FT REPEAT 2905 2942 PLECTIN 6.
 FT REPEAT 2943 2980 PLECTIN 7.
 FT REPEAT 2981 3018 PLECTIN 8.
 FT REPEAT 3019 3056 PLECTIN 9.
 FT REPEAT 3057 3094 PLECTIN 10.
 FT REPEAT 3274 3311 PLECTIN 11.
 FT REPEAT 3312 3349 PLECTIN 12.
 FT REPEAT 3350 3387 PLECTIN 13.

FT REPEAT 3388 3425 PLECTIN 14.
 FT REPEAT 3429 3463 PLECTIN 15.
 FT REPEAT 3609 3646 PLECTIN 16.
 FT REPEAT 3647 3684 PLECTIN 17.
 FT REPEAT 3685 3722 PLECTIN 18.
 FT REPEAT 3723 3760 PLECTIN 19.
 FT REPEAT 3764 3797 PLECTIN 20.
 FT REPEAT 3800 3834 PLECTIN 21.
 FT REPEAT 3852 3889 PLECTIN 22.
 FT REPEAT 3890 3927 PLECTIN 23.
 FT REPEAT 3928 3965 PLECTIN 24.
 FT REPEAT 3966 4003 PLECTIN 25.
 FT REPEAT 4007 4044 PLECTIN 26.
 FT REPEAT 4043 4094 PLECTIN 27.
 FT REPEAT 4197 4234 PLECTIN 28.
 FT REPEAT 4235 4272 PLECTIN 29.
 FT REPEAT 4273 4310 PLECTIN 30.
 FT REPEAT 4311 4348 PLECTIN 31.
 FT REPEAT 4349 4386 PLECTIN 32.
 FT DOMAIN 4039 4089 BINDING TO INTERMEDIATE FILAMENTS (BY SIMILARITY).
 FT MOD RES 4414 4429 4 X 4 AA TANDEM REPEATS OF G-S-R-X.
 FT SEQUENCE 4328 4328 PHOSPHORYLATION (BY CDC2).
 SQ SEQUENCE 4473 AA; 509015 MW; E14615D361E3484 CRC64;
 Query Match 6.3%; Score 114.5; DB 1; Length 4473;
 Best Local Similarity 19.6%; Pred. No. 20;
 Matches 73; Conservative 66; Mismatches 172; Indels 61; Gaps 10;
 QY 8 ECGHMLAEQFKOKAL-----RPAVIDREVL-----YQFDEMEESHLPYI 49
 DB 1906 EORREAEERVOXSLAEERAROKALEVEERLAKKVEBARRLRERAEQESRQLOLA 1965
 QY 50 ECGHMLAEQFKOKALAFASRGNHVSPTHYVPESDAPPAIIPDEVLHREFDEMEAFASR 109
 DB 1966 QEAQAQKRLQAEKKAHAFVQOR-----EEELQOTLQEQSMLERLGAEAA-RRA 2015
 QY 110 GNVSPHYVPESDASQAPYIEQAVIAH-----QFKKVLAFASRGHSDPTHYVP 163
 DB 2016 AEAEAEAREQAEEREAQSRQVEAEERLKQSAEQAOARQAQAAAEKLEKAEQEMARR 2075
 QY 164 SDAKPAIIPDEVLHREFDEMEESHLPYIECGHMLAEQFKOKALAFASRGNHVAIPHVY 223
 DB 2076 AAEQALAKKQQA-----PAEMEKHKKFAEQTLROKAQVQEQETLRLQ-----L 2121
 QY 224 TESDAKPAIIPDEVLHREFDEMEESHLPYIECGHMLAEQFKOKALAFASRGNHVAIPHVY 283
 DB 2122 EETDHOQSIL--DEELQRLKAEVTEARQSO-----VEELFVVRVQMEELGKTKA 2171
 DB 284 QFKKVLAFASRGNHVSPTHYVPESDAPPAIIPDEVLHREFDEMEAFASRGNHVAIPHVY 340
 QY 2172 RIAEAEALILRLDKNTQRFLEEAERKQVAEAEALSVAAQEAARLRQLAEDDLAQOR 2231
 QY 341 QRAEMLKSKIO 352
 DB 2232 ALAEKMLKEXMO 2243
 RESULT 13
 AAC4 HUMAN STANDARD; PRT; 911 AA.
 ID AAC4_HUMAN
 AC Q43707; 076048;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Alpha-actinin 4 (Non-muscle alpha-actinin 4) (F-actin cross linking protein).
 GN ACTN4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxId=9606;
 RN [1]

RP SEQUENCE OF 4-911 FROM N.A.
 RX MEDLINE=9817716; PubMed=9508771;
 RX Honda K., Yamada T., Endo R., Ino Y., Gotoh M., Tsuda H., Yamada Y.,
 RA Chiba H., Hirohashi S.;
 RT "Actinin-4, a novel actin-binding protein associated with cell
 RT motility and cancer invasion.";
 RL J. Cell Biol. 140:1383-1393(1998).
 RN [2]
 RP ERRATUM.
 RA Honda K., Yamada T., Endo R., Ino Y., Gotoh M., Tsuda H., Yamada Y.,
 RA Chiba H., Hirohashi S.;
 RL J. Cell Biol. 143:276-276(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Neuroblastoma;
 RX MEDLINE=20120478; PubMed=10656685;
 RA Nikolopoulos S.N., Spengler B.A., Kisselbach K., Evans A.E.,
 RA Biedler J.L., Ross R.A.;
 RT "The human non-muscle alpha-actinin protein encoded by the ACTN4 gene
 RT suppresses tumorigenicity of human neuroblastoma cells.";
 RL Oncogene 19:380-386(2000).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshlyuk S., Carninci P., Prange C.,
 RA Bask S.S., Loquellano N.A., Peters G.J., Abriemson R.D., Mulhany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs S.W.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.J., Skalska U., Smalls D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Maira M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [5]
 RP SEQUENCE OF 1-170 FROM N.A.
 RA Ottewill B., Obermaier B., Mewes H.-W., Gassenhuber J.,
 RA Wiemann S.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE OF 1-218 FROM N.A.
 RA Isogai T., Otsubo T., Sugiyama T.;
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP VARIANTS FSGS1 GIU-255; ILE-259 AND PRO-262.
 RC TISSUE=Lymphocytes;
 RX MEDLINE=20164321; PubMed=10700177;
 RA Kaplan J.M., Kim S.H., North K.N., Renke H., Correia L.A.,
 RA Tong H.-Q., Mathis B.J., Rodriguez-Perez J.-C., Allen P.G.,
 RA Beggs A.H., Pollak M.R.;
 RT "Mutations in ACTN4, encoding alpha-actinin-4, cause familial focal
 RT segmental glomerulosclerosis.";
 RL Nat. Genet. 24:251-256(2000).
 CC -!- FUNCTION: F-ACTIN CROSS-LINKING PROTEIN WHICH IS THOUGHT TO ANCHOR
 CC ACTIN TO A VARIETY OF INTRACELLULAR STRUCTURES. THIS IS A BUNDLING
 CC PROTEIN.
 CC -!- SUBUNIT: HOMODIMER, ANTIPARALLEL (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: NUCLEAR AND CYTOSOLIC. COLOCALIZES WITH
 CC ACTIN STRESS FIBERS. NUCLEAR TRANSLLOCATION CAN BE INDUCED BY THE
 CC PI3 KINASE INHIBITOR MORTALININ OR BY CYTOCHALASIN D. EXCLUSIVELY
 CC LOCALIZED IN THE NUCLEUS IN A LIMITED NUMBER OF CELL LINES (BREAST
 CC CANCER CELL LINE MCF7, ORAL FLUOR CANCER IMC2, AND BLADDER CANCER

Query Match	Best Local Similarity	6.2%	Score 113.5	DB 1	Length 911
Matches	79	Conservative	66	Mismatches	132
				Indels	117
				Gaps	19
QY	7	LEQGMLAEQFKOKA-----	-----LRPAVIPDREVLYQGFDEMEE--	ASHL	46
DB	408	LERLDHLAEKFRQASISHEAMTDGKEMLKHRIYETATLSLDIKALIRKHEAFESDLAAHQ	467		
QY	47	PYIQGQQLAEQFKOKALAFASRONHVSPTHYIPES--DARRPAIIPD--REVLYHR-----	E	98	
DB	468	DRVQVQIAIAIQELNE--LDYYDSHNVTTRQCKICDQWDAGSLTHSRREALETEKOLEA	525		
QY	99	FDEMEAFASRGNNHVSDAHVPRESDAQAPYIEQACVIAHQPEKVLAFASRGNNDSPT	156		
DB	526	IDQHLLEYAKRA-----APFNMMESAMEDLDQMFIYHTT--EELEGILSA--HDOFK	574		
QY	159	HYVEESD--AKRAIIPDREVLYREFDEMEESQHL-----PY-----	195		
DB	575	STLPDADREBEAL-----AIHKKEQRAIAEENHIKLSGNSYTTVTTPQIINSKWEKVOQLV	630		
QY	196	-----GNNLAEQFKOKA-----LAFASGNNHVP-----	THY	222	
DB	631	PKRDHALLEESQKOSNEHLRROFASQANVANGWIGTKEEIGRISIEGNGTLEDOLSHL	690		
QY	223	VTESDAKRALVPDKEVLYQGYDMEEEAFASRGNNHVPATHVVE-----	SDASASL	272	
DB	691	KQYERSIVDYKPNLDLEQGHQHLQELIIPDNKH---TNYTMEHIRGVGEQLLTIAARTI	747		
QY	273	PYWDE---TRAIAGQFEKVLAFASRGNNHVSPPH	303		
DB	748	NEVENQILTRDAKCISQEQWQEFASRNHHDKDH	781		

RESULT 14

PLE1_HUMAN

ID _PLE1_HUMAN STANDARD; PRT; 4684 AA.

AC Q15143: Q15148; Q16640;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Plectin 1 (PLTN) (PCN) (Hemidesmosomal protein 1) (HD1).

GN PLEC1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

NCBI_Taxid=9606;

NN [1]


```

FT REPEAT 4484 4521 PLECTIN 31.
FT REPEAT 4522 4559 PLECTIN 32.
FT REPEAT 4560 4597 PLECTIN 33.
FT DOMAIN 4250 4300 BINDING TO INTERMEDIATE FILAMENTS
(BY SIMILARITY).
FT DOMAIN 4625 4640 4 X 4 AA TANDEN REPEATS OF G-S-R-X.
FT MOD_RES 4539 4539 PHOSPHORYLATION (BY CDC2)
(BY SIMILARITY).
FT VARSPLIC 1 174 MWAGMLRDLRAIYVLFREGVAVAKRRPSLHPHV
CVTNLQWRAMASLARGVYRTFAMCHFFVLTGTAHL
ROYLHPPEIVASLQVRPVAMPARTPTVAVGGL
GSPKRGPLTEBQRLYRKLELVSPETVPATQTLA
RPGEPAPAT -> MSGDAEAVRAVSEVNSGSPSPGD
TLPNLIKTRSRSGGAGSGVSLDPAERAVIRIA (in
isoform 2 and isoform 3).
FT /FTid=VSP 005030.
Missing (in isoform 3).
FT VARSPLIC 409 412 /FTid=VSP 005031.
Missing (in MD-EBS).
FT VARIANT 429 429 L -> LL (in MD-EBS).
/FTid=VAR 011316.
Missing (in MD-EBS).
FT VARIANT 1003 1005 /FTid=VAR 011317.
S -> F (IN REF. 2).
FT CONFLICT 185 185 N -> D (IN REF. 2).
FT CONFLICT 259 259 H -> N (IN REF. 2).
FT CONFLICT 550 550 I -> V (IN REF. 2).
FT CONFLICT 560 560 Q -> R (IN REF. 2).
FT CONFLICT 706 706 N -> Y (IN REF. 2).
FT CONFLICT 886 886 V -> A (IN REF. 2).
FT CONFLICT 1002 1002 V -> L (IN REF. 2).
FT CONFLICT 1309 1309 L -> V (IN REF. 2).
FT CONFLICT 1321 1321 V -> L (IN REF. 2).
FT CONFLICT 1334 1334 I -> M (IN REF. 2).
FT CONFLICT 1534 1534 A -> T (IN REF. 2).
FT CONFLICT 1662 1662 MLC -> RLR (IN REF. 2).
FT CONFLICT 1688 1690 Q -> E (IN REF. 2).
FT CONFLICT 1767 1767 L -> R (IN REF. 2).
FT CONFLICT 1789 1789 K -> N (IN REF. 2).
FT CONFLICT 1910 1910 S -> R (IN REF. 2).
FT CONFLICT 2154 2154 S -> Q (IN REF. 2).
FT CONFLICT 2160 2161 R -> A (IN REF. 2).
FT CONFLICT 2215 2215 S -> A (IN REF. 2).
FT CONFLICT 2244 2244

```

Query Match 6.2%; Score 113.5; DB 1; Length 4684;

Best Local Similarity 19.7%; Pred. No. 24; Mismatches 176; Indels 51; Gaps 8;

Matches 71; Conservative 63; Mismatches 176; Indels 51; Gaps 8;

```

14 AEQFKOKAL-----RPAVTPREV-----YQEFDEMEASHLPYIEQGMQL 55
2123 AEERQKSLAEAEARQKALVEVERLKAKVEARSUREABESARQLOLAQEAOK 2182
56 AEQFKOKALFASRGNHVSPTHYVESDAPPAIIPDREVLHREFDEMEAFASRGNHVS 115
2183 RLQAEKKAHAFVQOK-----EOELQOTLQEOSVLDRLRGEAFA--RRAAEFAE 2232
QY 116 AHYVESASQAPRTIEQAOVIAHFKKVLAFASRGNHVSPTHYVESDAKPAIIPDR 174
DB 2233 ARVQAEEREAQSRROVEEERLKQSAEEOQAQAAAEKLRKAEQEAARRQQAQ 2292
QY 175 EVLYREFDEMEESQHLPYIEQGMMLAEQFKOKALFASRGNHVAPTHYVESDAKPA 234
DB 2233 ALRQQAADAEAEKKAHFAEQTLRQKAVQEOELTLRLQ-----LEETDHQKML- 2342
QY 235 DKEVLYQOYDEMEEAFAASRGNHVAPTHYVESDASASLPYMETRAIAGQFKKVLAF 294
DB 2343 -----DEELQRLKAEATEARQSQVEEELFSVRQVMEELSKLXARIEAENRALIL 2393
QY 235 RGNHVSPTHYVESDAPPAIIPDREVLHREFDEMEAFASRGNHVSPTHYVESDA 351
DB 2394 RDKONTORFLQEAEEKKQVAAEAAARLSVAQEAARLRQLAEEDLAQAPALAEKML 2453
QY 352 Q 352

```

```

DB 2454 Q 2454
RESULT 15
PTE1_RAT STANDARD; PRT; 4687 AA.
AC P30427; O08879; O08880; O08881;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Plectin 1 (PLTN) (PCN).
DN PLECI.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CX NCBI_TaxId=10116;
RN (1)
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Glia1 tumor;
RX MEDLINE=9126156; PubMed=2050743;
RA Wiche G., Becker B., Luber K., Weitzer G., Castanon M.J.,
RT Hauptmann R., Stratawa C., Stewart M.;
RT "Cloning and sequencing of rat plectin indicates a 466-kD polypeptide
RT chain with a three-domain structure based on a central alpha-helical
RT coiled coil."
RL J. Cell Biol. 114:83-99(1991).
RN (2)
RP REVISIONS.
RC TISSUE=Glia1 tumor;
RX MEDLINE=96210632; PubMed=8633055;
RA Liu C.-G., Maercker C., Castanon M.J., Hauptmann R., Wiche G.;
RT "Human plectin: organization of the gene, sequence analysis, and
RT chromosome localization (6q24).";
RL Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283(1996).
RN (3)
RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 2, 3 AND 4), AND TISSUE
RP SPECIFICITY.
RC TISSUE=Glia1 tumor;
RX MEDLINE=97321050; PubMed=9177781;
RA Eliiott C.E., Becker B., Oehler S., Castanon M.J., Hauptmann R.,
RA Wiche G.;
RT "Plectin transcript diversity: identification and tissue distribution
RT of variants with distinct first coding exons and rodless isoforms."
RL Genomics 42:115-125(1997).
CC -!- FUNCTION: INTERLINKS INTERMEDIATE FILAMENTS WITH MICROTUBULES AND
CC MICROFILAMENTS AND ANCHORS INTERMEDIATE FILAMENTS TO DESMOSOMES OR
CC HEMIDESMOSOMES. MAY BE INVOLVED NOT ONLY IN THE CROSSLINKING AND
CC STABILIZATION OF CYTOSKELETAL INTERMEDIATE FILAMENTS NETWORK, BUT
CC ALSO IN THE REGULATION OF THEIR DYNAMICS.
CC -!- SUBUNIT: HOMODIMER OR HOMOTETRAMER.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=1;
CC IsoId=P30427-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P30427-2; Sequence=VSP_005050;
CC Name=3;
CC IsoId=P30427-3; Sequence=VSP_005051;
CC Name=4;
CC IsoId=P30427-4; Sequence=VSP_005052;
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST EXPRESSION IN
CC SKELETAL MUSCLE AND LOWEST IN THYMUS.
CC -!- DOMAIN: THE N-TERMINUS INTERACTS WITH ACTIN, THE C-TERMINUS WITH
CC VIMENTIN, DESMIN, GFAP, CYTOKERATINS, LAMIN B; WHEREAS BOTH THE N-
CC AND THE C-TERMINUS CAN BIND INTEGRIN, BETA-4.
CC -!- PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM
CC INTERMEDIATE FILAMENTS DURING MITOSIS (BY SIMILARITY).
CC -!- SIMILARITY: Contains 1 actin-binding domain.
CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.
CC -!- SIMILARITY: Contains 3 plectin repeats.
CC -!- SIMILARITY: Contains 4 spectrin repeats.
CC -!- SIMILARITY: Belongs to the plectin or cytolinker family.
CC -!- CAUTION: ISOFORM 4 IS A FRAGMENT AT THE N-TERMINUS.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X59601; CAA42169.1; -
DR EMBL; U96274; AAC53209.1; -
DR EMBL; U96275; AAC53210.1; -
DR EMBL; U96276; AAC53211.1; -
DR PIR; A39638; A39638.
DR HSSP; Q01082; 1BKR.
DR InterPro; IPR001589; Actbind_accnin.
DR InterPro; IPR001715; Calponin-like.
DR InterPro; IPR001101; Plectin_repeat.
DR InterPro; IPR005328; S10_plectin.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00307; CH_2.
DR Pfam; PF00681; Plectin_21.
DR Pfam; PF03501; S10_plectin_1.
DR ProDom; PD006662; S10_plectin_N; 1.
DR SMART; SM00033; CH_2.
DR SMART; SM00250; PLEC; 33.
DR SMART; SM00150; SPEC; 33.
DR PROSITE; PS00019; ACTININ_1; FALSE_NEG.
DR PROSITE; PS00020; ACTININ_2; FALSE_NEG.
DR PROSITE; PS50021; CH_2.
DR Coiled coil; Repeat: Structural protein; Cytoskeleton; Actin-binding;
KM Phosphorylation; Alternative splicing.
KM Phosphorylation; Alternative splicing.
FT DOMAIN 1 1473 GLOBULAR 1.
FT DOMAIN 1474 2758 CENTRAL FIBROUS ROD DOMAIN.
FT DOMAIN 2759 4687 GLOBULAR 2.
FT DOMAIN 181 406 ACTIN-BINDING.
FT DOMAIN 185 288 CH 1.
FT DOMAIN 301 403 CH 2.
FT REPEAT 648 722 SPECTRIN 1.
FT REPEAT 743 827 SPECTRIN 2.
FT REPEAT 840 933 SPECTRIN 3.
FT REPEAT 1318 1418 SPECTRIN 4.
FT DOMAIN 1472 1692 COILED COIL.
FT DOMAIN 1724 2760 COILED COIL (POTENTIAL).
FT REPEAT 2791 2828 COILED COIL (POTENTIAL).
FT REPEAT 2829 2866 PLECTIN 1.
FT REPEAT 2867 2904 PLECTIN 2.
FT REPEAT 2905 2942 PLECTIN 3.
FT REPEAT 2943 2980 PLECTIN 4.
FT REPEAT 2984 3018 PLECTIN 5.
FT REPEAT 3119 3156 PLECTIN 6.
FT REPEAT 3157 3194 PLECTIN 7.
FT REPEAT 3195 3232 PLECTIN 8.
FT REPEAT 3233 3270 PLECTIN 9.
FT REPEAT 3271 3308 PLECTIN 10.
FT REPEAT 3311 3346 PLECTIN 11.
FT REPEAT 3488 3525 PLECTIN 12.
FT REPEAT 3526 3563 PLECTIN 13.
FT REPEAT 3564 3601 PLECTIN 14.
FT REPEAT 3602 3639 PLECTIN 15.
FT REPEAT 3643 3677 PLECTIN 16.
FT REPEAT 3823 3860 PLECTIN 17.
FT REPEAT 3861 3898 PLECTIN 18.
FT REPEAT 3899 3936 PLECTIN 19.
FT REPEAT 3937 3974 PLECTIN 20.
FT REPEAT 3978 4011 PLECTIN 21.
FT REPEAT 4066 4103 PLECTIN 22.
FT REPEAT 4104 4141 PLECTIN 23.
FT REPEAT 4142 4179 PLECTIN 24.
FT REPEAT 4180 4217 PLECTIN 25.
FT REPEAT 4221 4255 PLECTIN 26.
FT REPEAT 4268 4308 PLECTIN 27.
FT REPEAT 4411 4448 PLECTIN 28.
FT REPEAT 4411 4448 PLECTIN 29.

FT REPEAT 4449 4486 PLECTIN 30.
FT REPEAT 4487 4524 PLECTIN 31.
FT REPEAT 4525 4562 PLECTIN 32.
FT REPEAT 4563 4600 PLECTIN 33.
FT DOMAIN 4253 4303 BINDING TO INTERMEDIATE FILAMENTS.
FT DOMAIN 4628 4643 4 X 4 AA TANDEN REPEATS OF G-S-R-X.
FT MOD_RES 4542 4542 PHOSPHORYLATION (BY CDC2)
FT VARSPPLIC 1 180 (BY SIMILARITY).
FT VARSPPLIC 1 180 MWAGMLMPLDQLRAIYEVLPREGVWVAKKRRPSLHPHP
FT VARSPPLIC 1 180 GVTNLQVNRAMTSLKARGLVRETFAMCHFYVLTNEGIDHL
FT VARSPPLIC 1 180 ROYLHLPEIYVPSLQVRVRVYVAMVPPARRSPHYQTQGP
FT VARSPPLIC 1 180 LGCPKRGPLPAEDPAREREQVYRKRKEEGCAPETPVVSAT
FT VARSPPLIC 1 180 IGVTLARGPPEPTPAT -> MSQQLRPPPEBGSSKXTSS
FT VARSPPLIC 1 180 EDNLYLAVLRASGKK (in isoform 2).
FT VARSPPLIC 1 180 /FTId=VSP_005050.
FT VARSPPLIC 1 180 MWAGMLMPLDQLRAIYEVLPREGVWVAKKRRPSLHPHP
FT VARSPPLIC 1 180 GVTNLQVNRAMTSLKARGLVRETFAMCHFYVLTNEGIDHL
FT VARSPPLIC 1 180 ROYLHLPEIYVPSLQVRVRVYVAMVPPARRSPHYQTQGP
FT VARSPPLIC 1 180 LGCPKRGPLPAEDPAREREQVYRKRKEEGCAPETPVVSAT
FT VARSPPLIC 1 180 IGVTLARGPPEPTPAT -> MEPSGSLFSLVVGHVLSLA
FT VARSPPLIC 1 180 AVWHRKGRQAODEQ (in isoform 3).
FT VARSPPLIC 1 180 /FTId=VSP_005051.
FT VARSPPLIC 1 180 MWAGMLMPLDQLRAIYEVLPREGVWVAKKRRPSLHPHP
FT VARSPPLIC 1 180 GVTNLQVNRAMTSLKARGLVRETFAMCHFYVLTNEGIDHL
FT VARSPPLIC 1 180 ROYLHLPEIYVPSLQVRVRVYVAMVPPARRSPHYQTQGP
FT VARSPPLIC 1 180 LGCPKRGPLPAEDPAREREQVYRKRKEEGCAPETPVVSAT
FT VARSPPLIC 1 180 IGVTLARGPPEPTPAT -> DVNSGSSGSPSGDITPVNLG
FT VARSPPLIC 1 180 KTORRRRSGSGVSGVGLDPAERAVIRIA (in
FT VARSPPLIC 1 180 isoform 4).
FT VARSPPLIC 1 180 /FTId=VSP_005052.
FT VARSPPLIC 1 180 /FTId=VSP_005053.
SQ SEQUENCE 4687 AA; 533527 MW; 9966C871B929751 CR664;
Query Match 6.0%; Score 109.5; DB 1; Length 4687;
Best Local Similarity 18.8%; Pred. No. 44;
Matches 70; Conservative 67; Mismatches 174; Indels 61; Gaps 10;
QY 8 EQGMLAQFQKAL--RPAVIPREVLYQFDP-----EMEESHLPYI 49
DB 2120 EQRREAEERVORSLAEEDARORKVALEVERLKAKEBARLRERAEQESARQLA 2179
QY 50 EQGMQLAQFQKALAFASRONHVSPTHYVESAPRAIPDREVLHREFDEMEAFASR 109
DB 2180 QEAQKRLQAEKKAHAFVQOR-----EEELQOTLQOEQMLERLRSEADA--RRA 2229
QY 110 GNHVSFAHYVESDASQAPYIEQAOVIAHQKE-----KVLAFAKGNHDSPTHYVE 163
DB 2230 AEEAEAEAEQAEEREAQKQVEAEERLKQSAEEQAOQAQAAAEKLRKAEAEQAEARR 2289
QY 164 SDAKPAIIPDREVLHREFDEMEESQHLPIYIEQGMQLAQFQKALAFASRONHVA 223
DB 2290 AQAEDQALAKQQA-----DAEMEKHKKFAEOTLRQKAOVEOELTTLRLQ-----L 2335
QY 224 TESDAKPAIIPDREVLHREFDEMEAFASRGNHVAAPHYVESDASASLPVMDETRAIAG 283
DB 2336 EETDQKSL--DELDQRLKAETVBAARQSQ-----VEELFSRVVMEELGKKA 2385
QY 284 QFKEKVLAFASRGNHVSBRHYVPESEPDVVTPDKIELYEADE---MEASKALIEEG 340
DB 2386 RIEAENRLLILRDQNTORFLSEAEKMKKQVAEEAARLSVAQGAARLRQAEDDLAQOR 2445
QY 341 QRMAMKLSKIQ 352
DB 2446 ALAEKMLKEKMQ 2457

Search completed: July 25, 2003, 13:10:14
Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: July 25, 2003, 13:06:36 ; Search time 41 Seconds

(without alignments)
825.643 Million cell updates/sec

Title: US-09-491-146A-52

Perfect score: 1818

Sequence: 1 AAHPLYEQGMHLAEQFKOK.....KALIEEGQMAEMLSKIQ 352

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

11 number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	263	14.5	3010	1	S18030 genome polyprotein
2	260	14.3	3010	1	GNMVCU genome polyprotein
3	259	14.2	3011	1	S40770 genome polyprotein
4	258	14.2	3010	1	GNMVTM genome polyprotein
5	256	14.1	492	2	PS0326 polyprotein - hepa
6	256	14.1	3010	1	GNMVTM genome polyprotein
7	256	14.1	3010	1	A45573 genome polyprotein
8	254	14.0	3011	1	GNMVC3 genome polyprotein
9	247	13.6	3011	1	GNMVC3 genome polyprotein
10	207	11.4	716	2	J01366 polyprotein - hepa
11	203.5	11.2	3033	1	J01303 genome polyprotein
12	201.5	11.1	3033	1	GNMVT8 genome polyprotein
13	195	10.7	3014	1	UC5620 genome polyprotein
14	188.5	10.4	876	1	PC2219 polypeptide - hepa
15	188	10.3	125	2	S35629 hypothetical prote
16	176.5	9.7	209	2	PC1306 genome polyprotein
17	173.5	9.5	142	2	PC1307 genome polyprotein
18	164.5	8.5	411	2	S47436 flagellar antigen
19	121.5	6.7	911	2	UC7186 alpha-actinin-4 -
20	114.5	6.3	1621	2	A62255 hypothetical prote
21	114.5	6.3	5327	2	T13564 microtubule-associ
22	113.5	6.2	1224	2	T14007 microtubule-associ
23	113.5	6.2	1262	2	T22523 microtubule-associ
24	113.5	6.2	1778	2	A61116 hypothetical prote
25	113.5	6.2	4574	2	G02520 internalin protein
26	113.5	6.2	4684	2	A59404 plectin (human
27	113.5	6.2	5170	2	T15348 plectin (imported)
28	113	6.2	7962	2	T18346 hypothetical prote
29	112.5	6.2	1385	2	T21706 elastic titin - hu
					hypothetical prote

30	112.5	6.2	1929	2	T21559 hypothetical prote
31	112	6.2	1785	2	T21558 hypothetical prote
32	110.5	6.1	506	2	S47439 I2 protein - trypa
33	109.5	6.0	1948	2	S00485 gene 11-1 protein
34	109.5	6.0	4687	1	A39638 plectin - rat
35	107.5	5.9	1733	1	RNB42L DNA-directed RNA p
36	107.5	5.9	5762	2	A41819 prolone-rich pepi
37	107	5.9	910	1	S73361 dnal homolog prote
38	107	5.9	2957	2	T33152 hypothetical prote
39	106.5	5.9	723	2	H85092 hypothetical prote
40	106.5	5.9	1145	2	T33606 hypothetical prote
41	106.5	5.9	1736	2	A47747 tight junction pro
42	106	5.8	948	2	S75991 hypothetical prote
43	106	5.8	2748	2	S57976 nuclear migration
44	104.5	5.7	327	2	S31781 HRP1 protein - ma
45	104.5	5.7	419	2	G75062 probable flagella-

ALIGNMENTS

RESULT 1

S18030 genome polyprotein - hepatitis C virus (isolate JKI)

N:Contains: capsid protein C; envelope protein M; hepatitis C virus (EC 3.4.21.98) (nonstr)

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus

A:Variety: isolate JKI

C:Date: 19-May-2000 #sequence revision 19-May-2000 #text_change 23-Mar-2001

C:Accession: S18030; S33570; A48332; S18029

R:Honda, M.; Kaneko, S.; Masehara, U.; Kobayashi, K.; Murakami, S.

submitted to the EMBL Data Library, September 1991

A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single i

A:Reference number: S18028

A:Accession: S18030

A:Molecule type: genomic RNA

A:Residues: 1-310 <HON>

A:Cross-references: EMBL:X61596; NID:G59478; PIRN:CAA4793.1; PID:G59479

A:Experimental source: isolate JKI from an individual

R:Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.

Arch. Virol. 128, 163-169, 1993

A:Title: Sequence analysis of putative structural regions of hepatitis C virus isola

A:Reference number: A48332; MUID:91119270; PMID:8380322

A:Accession: S33570

A:Molecule type: genomic RNA

A:Residues: 1-547, 'T', '549-621', 'V', '623-624', 'S', '626-652', 'DL', '655-761', 'T', '763-782' <HON>

A:Cross-references: EMBL:X61591

A:Note: this sequence is inconsistent with the nucleotide translation

as Trp, and Trp for residue 771 as Ser

A:Note: sequence extracted from NCBI backbone (NCBI:121747, NCBI:121748)

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; s

F:2-115/Product: capsid protein C #status predicted <CP>

F:116-191/Product: envelope protein M #status predicted

F:192-389/Product: major envelope protein E #status predicted <ME>

F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>

F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F:1007-1615/Product: nonstructural protein NS3 #status predicted <NS3>

F:1620-1237/Product: nucleocapsid protein NS4 #status predicted <NS4>

F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEHH motif

F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>

F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

Query Match 14.5% Score 263; DB 1; Length 3010;

Best Local Similarity 32.5% Pred. No. 3.1e-09;

Matches 87; Conservativity 27; Mismatches 74; Indels 80; Gaps 11;

Ox 23 RPAITPDEVLYEQFDEME--ASHLPYIEQGMHLAEQFKOKALAF--ASRONHVSPTHY 78

Db 1746 -----AAP-----VSSKRALETFWAKHMMNISCIGYIAGSLTLPGNPAIASLMAF 1793
QY 134 -AOVIAHQFKEKVLAF-----ASRGNHDSPTHYVPSDAKPAI-----170
Db 1794 TASITSPLTOSTLFTNLIGVVAQAOLAPSAASAFVAGIAGAAGVSGIGKVLVDILA 1853
QY 171 -----IPDREVLVREFDEMESQHL--PYIEQGMML-----AE 201
Db 1854 GYGAGVAGALVAFKVMSEMPSTEDLVNLLPALISPGALVGVVCAILRRHVGPGEAV 1913
QY 202 QFKOKALAFASRGNHVAPTHYVTESDA 228
Db 1914 QMNRRLIAFASRGNHVSPTHYVPSDA 1940

RESULT 7
A:Accession: A45573
Genome polyprotein - hepatitis C virus (strain JT)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome from a single Japanese carrier: S
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
C:Accession: A45573
R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata, A.;
Virus Res. 23, 39-53, 1992
A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: S
A:Reference number: A45573; PMID:92295714; PMID:1318627
A:Accession: A45573
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3010 <TRAN>
A:Cross-references: GB:D11168; GB:D01171; NID:g221612; PIDN:BA01943.1; PID:g221613
A:Experimental source: HCV-JT
A:Note: sequence extracted from NCBI backbone (NCBI:106206, NCBI:P.106207)
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:132-389/Product: major envelope protein E #status predicted <ME>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEKH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>
014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 14.1%; Score 256; DB 1; Length 3010;
Best Local Similarity 31.5%; Pred. NO. 9e-09;
Matches 84; Conservative 31; Mismatches 74; Indels 78; Gaps 10;

QY 23 RPAAVDPREVLYOEFDMEEA--ASHLPYIEQGMQLAEQFKOKALAF---ASRGNHVSPTHY 78
Db 1691 RPAAVDPREVLYREFDEMEECSHLPYIEQGMQLAEQFKOKALGLQTSRAEV-----1745
QY 79 VPESDARPAIIPDREVLHREFDEMEAFASR--GNHVSPTHYVPSDASQAAPYIEQ---133
Db 1743 ---AEEAAPVVSER-----WRALFAFWAGMMNFIISGIOYLAGSLTPGNPAIASLMAF 1793
QY 134 -AOVIAHQFKEKVLAF-----ASRGNHDSPTHYVPSDAKPAI-----170
Db 1794 TASITSPLTOSTLFTNLIGVVAQAOLAPSAASAFVAGIAGAAGVSGIGKVLVDILA 1853
QY 171 -----IPDREVLVREFDEMESQHL--PYIEQGMML-----AE 201
Db 1854 GYGAGVAGALVAFKVMSEMPSTEDLVNLLPALISPGALVGVVCAILRRHVGPGEAV 1913
QY 202 QFKOKALAFASRGNHVAPTHYVTESDA 228
Db 1914 QMNRRLIAFASRGNHVSPTHYVPSDA 1940

RESULT 8
GNMVC3
Genome polyprotein - hepatitis C virus (strain HCV-1)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome from a single Japanese carrier: S
C:Species: hepatitis C virus
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001
C:Accession: A39166; PMID:92268871; PMID:1316939
R:Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.
Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991
A:Title: Genetic organization and diversity of the hepatitis C virus.
A:Reference number: A39166; PMID:91172826; PMID:1848704
A:Accession: A39166
A:Molecule type: mRNA
A:Residues: 1-3011 <CHO>
A:Cross-references: GB:M62321; NID:g329873; PIDN:AAA5676.1; PID:g329874
R:Chan, S.M.; McMahon, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap
J. Gen. Virol. 73, 1131-1141, 1992
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship
A:Reference number: P00403; PMID:92268871; PMID:1316939
A:Accession: P00403
A:Molecule type: genomic RNA
A:Residues: 1577-1633 <CH>
A:Cross-references: DDBJ:D10128
A:Experimental source: isolates E-b16
A:Accession: P00404
A:Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1577-1633 <CH2>
A:Experimental source: isolates E-b17
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstru
F:115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:132-389/Product: major envelope protein E #status predicted <ME>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEKH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,305,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,20

Query Match 14.0%; Score 254; DB 1; Length 3011;
Best Local Similarity 31.8%; Pred. NO. 1.2e-08;
Matches 85; Conservative 27; Mismatches 77; Indels 78; Gaps 10;

QY 23 RPAAVDPREVLYOEFDMEEA--HLPYIEQGMQLAEQFKOKALAF---ASRGNHVSPTHY 78
Db 1691 RPAAVDPREVLYREFDEMEECSHLPYIEQGMQLAEQFKOKALGLQTSRAEV-----1745
QY 79 VPESDARPAIIPDREVLHREFDEMEAFASR--GNHVSPTHYVPSDASQAAPYIEQ---133
Db 1746 -----IAPAV-----QTNMOKLETFWAKHMMNISCIGYIAGSLTLPGNPAIASLMAF 1793
QY 134 -AOVIAHQFKEKVLAF-----ASRGNHDSPTHYVPSDAKPAI-----170
Db 1794 TAAVTSPLTOSTLFTNLIGVVAQAOLAPSAASAFVAGIAGAAGVSGIGKVLVDILA 1853
QY 171 -----IPDREVLVREFDEMESQHL--PYIEQGMML-----AE 201
Db 1854 GYGAGVAGALVAFKVMSEMPSTEDLVNLLPALISPGALVGVVCAILRRHVGPGEAV 1913
QY 202 QFKOKALAFASRGNHVAPTHYVTESDA 228
Db 1914 QMNRRLIAFASRGNHVSPTHYVPSDA 1940

RESULT 9

GNMVAH

genome polypeptide - hepatitis C virus (strain H)

N.Contains: capsid protein C; envelope protein M; hepatitisin (EC 3.4.21.98) (nonstructu

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C.Species: hepatitis C virus

A.Note: host Homo sapiens (man)

C.Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001

A.Accession: A36814; A41546

R.Inchoupe, G.; Zebede, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.

Submitted to GenBank, July 1992

A.Description: Genomic structure of the human prototype strain H of hepatitis C virus; C

A.Reference number: A36814

A.Accession: A36814

A.Molecule type: genomic RNA

A.Residues: 1-3011 <INC>

A.Cross-references: GB:M67463, NID:G229737, PIDN:AAA5534.1, PID:G229738

R.Inchoupe, G.; Zebede, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.

Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991

Title: Genomic structure of the human prototype strain H of hepatitis C virus; compari

Reference number: A41546, MUID:92052256, PMID:1658800

Contents: annotation

A>Note: neither amino acid nor nucleotide sequence is given

C:Superfamily: hepatitis C virus genome polypeptide

C:Keywords: ATP; capsid protein C; envelope protein; glycoprotein; hydrolase; nonstructura

F.1-115/Product: capsid protein C #status predicted <CPC>

F.116-191/Product: envelope protein M #status predicted <EPM>

F.192-389/Product: major envelope protein E #status predicted <MEP>

F.390-728/Product: nonstructural protein NS1 #status predicted <NS1>

F.730-1006/Product: nonstructural protein NS2 #status predicted <NS2>

F.1007-1615/Product: hepatitisin #status predicted <NS3>

F.1320-1337/Region: nucleotide-binding motif A (P-loop)

F.1312-1317/Region: nucleotide-binding motif B

F.1316-1319/Region: DEXH motif

F.1616-1662/Product: nonstructural protein NS4a #status predicted <N4A>

F.1867-2013/Product: nonstructural protein NS4b #status predicted <N4B>

F.2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

F.196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23

Query Match

Best Local Similarity 13.6%; Score 247; DB 1; Length 3011;

Matches 86; Conservative 26; Mismatches 77; Indels 78; Gaps 11;

23 RPAVDPREVLYREFDEMEAS-HLPYIEQGMALAEQFOKXALAF---ASRGNHVSPTHY 78

Db 1691 KPAIIPDREVLYREFDEMEASCHLPYIEQGMALAEQFOKXALGLLTQTSRAEV----- 1745

QV 79 VESDARPAIIPDREVLYREFDEMEASR-GNHVSPTHYVPSDASQAPYIEO--- 133

L 1746 -----ITPAV-----QTNMQLKLVFMAKHMNFISGIQYLAGLSTLPGNPAISMAF 1793

QY 134 -AOVIAHQEKVLAFA-----ASRGHDSPTHYVPSDAKPAI----- 170

Db 1794 TAAVVSPLTGTQTLFNLIGWVAQOLAPGAATAVGAGLGAALDSVGLKVLVDILA 1853

QY 171 -----IPDREVLYREFDEMEASCH-LPYI-----EQGMALAE----- 201

Db 1854 GYGAGVAGALVAFKISGEVSTEDLVNLPAILSPGALAVGVFASITRRVVGEGAV 1913

QY 202 QFOKALAFASRGNHVAPTHYTESDA 228

Db 1914 QMMNRLIAFASRGNHVSPTHYVESDA 1940

RESULT 10

JQ1366

polypeptide - hepatitis C virus (French isolate) (fragments)

C.Species: hepatitis C virus

C.Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000

C.Accession: JQ1366

R.Kremsdorf, D.; Potchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.

J. Gen. Virol. 72, 2557-2561, 1991

A>Title: Partial nucleotide sequence analysis of a French hepatitis C virus; implication

A:Reference number: JQ1366; MUID:92013977; PMID:1655961

A:Accession: JQ1366

A.Molecule type: genomic RNA

A.Residues: 1-716 <KRE>

C:Superfamily: hepatitis C virus genome polypeptide

C:Keywords: glycoprotein; polypeptide

F.84,90,97,115,143,199,223,243,290,312/Binding site: carbohydrate (Aen) (covalent)

Query Match 11.4%; Score 207; DB 2; Length 716;

Best Local Similarity 57.6%; Pred. No. 2.6e-06;

Matches 49; Conservative 4; Mismatches 16; Indels 16; Gaps 3;

QY 167 KPAIIPDREVLYREFDEMEASCHLPYIEQGMALAEQFOKXALAFASRGNHVAPTHYTE 225

Db 640 KPAIIPDREVLYREFDEMEASCHLPYIEQGMALAEQFOKXALGLLTQ-----TR 688

QY 226 SDKAPALVPDKEVLYQYDEMEAF 250

Db 689 SROAEVITPAVOTNMQR-----EAF 709

RESULT 11

JQ1303

genome polypeptide - hepatitis C virus (isolate HC-06)

N.Contains: capsid protein C; envelope protein M; hepatitisin (EC 3.4.21.98) (nonst

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C.Species: hepatitis C virus

C.Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 17-Nov-2000

C:Accession: JQ1303

R.Okamoto, H.; Okada, S.; Sugiyama, Y.; Kura, K.; Iizuka, H.; Machida, A.; Miyakaw

J. Gen. Virol. 72, 2697-2704, 1991

A>Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from

A:Reference number: JQ1303; MUID:92044440; PMID:1658196

A:Accession: JQ1303

A.Molecule type: genomic RNA

A.Residues: 1-3033 <OKA>

A:Cross-references: GB:D00944; NID:G221650; PIDN:BAA00792.1; PID:G221651

A:Experimental source: isolate HC-06 from a Japanese individual

C:Superfamily: hepatitis C virus genome polypeptide

C:Keywords: ATP; glycoprotein; hydrolase; P-loop; polypeptide; serine proteinase; c

F.2-115/Product: capsid protein C #status predicted <CPC>

F.116-191/Product: envelope protein M #status predicted <EPM>

F.192-389/Product: major envelope protein E #status predicted <MEP>

F.390-733/Product: nonstructural protein NS1 #status predicted <NS1>

F.734-1010/Product: nonstructural protein NS2 #status predicted <NS2>

F.1011-1619/Product: hepatitisin #status predicted <NS3>

F.1316-1321/Region: DEXH motif

F.1320-1323/Region: nucleotide-binding motif B

F.1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>

F.1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>

F.2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>

F.196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,20;

Query Match

Best Local Similarity 11.2%; Score 203.5; DB 1; Length 3033;

Matches 60; Conservative 10; Mismatches 37; Indels 29; Gaps 5;

QY 229 KPAVDPKEVLYQYDEMEAFASRGNHVAPTHYVESDAS---ASLPYMETRAIACOF 285

Db 1619 KPTLVGPTPLLYR-----LGSVTNEVTLTHPTKXATACQMDLVMSTWLACGV 1670

QY 286 KEXVIAFASRGN-----HVSPTHYVPSDASQAPYIEO-----ASKAAL 336

Db 1671 LAVAAYCLATGCACCTIIRLHVNR-----AVVAPDKEVLYEAFDEMEASRAAL 1721

QY 337 IEEGQMAEMLSKIQ 352

Db 1722 IEEGQRIEMLSKIQ 1737

RESULT 12

GNMV08

genome polypeptide - hepatitis C virus (strain HC-08)

N:Contains: capsid protein C; envelope protein M; hepatitisin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C>Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
A:Accession: A40250; P00397; P00559
R:Okamoto, H.; Kurai, K.; Okada, S.I.; Yamamoto, K.; Lizuka, H.; Tanaka, T.; Fukuda, S.;
Virology 188, 331-341, 1992
A:Title: Full-length sequence of a hepatitis C virus genome having poor homology to rep
A:Reference number: A40250; MUID:9220323; PMID:1314459
A:Accession: A40250
A:Molecule type: genomic RNA
A:Residues: 1-3033 <OKA>
A:Cross-references: GB:D10988; GB:D01221; NID:G221608; PIDN:BAA01761.1; PID:G221609
R:Chan, S.W.; McMahon, F.; Holmes, E.C.; Dow, B.; Peucheter, J.F.; Follett, E.; Yap, P.L.
J. Gen. Virol. 73, 1131-1141, 1992
A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
A:Reference number: P00393; MUID:92268871; PMID:1316939
A:Accession: P00397
A:Molecule type: genomic RNA
A:Residues: 2678-2754 <CHA>
A:Cross-references: DDBJ:D10134
A:Experimental source: isolate E-b12
R:Kato, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohnd
Biochem. Biophys. Res. Commun. 181, 279-285, 1991
A:Title: Distribution of plural HCV types in Japan.
A:Reference number: P00554; MUID:92068204; PMID:1720309
A:Accession: P00559
A:Molecule type: mRNA
A:Residues: 2678-2729 <KAT>
A:Cross-references: GB:D10562; GB:D90518; NID:G221523; PIDN:BAA01418.1; PID:G221524
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
F:115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EMP>
F:192-189/Product: major envelope protein E #status predicted <MEE>
F:790-733/Product: nonstructural protein NS1 #status predicted <NS1>
F:794-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F:1011-1619/Product: hepatitisin #status predicted <NS3>
F:1234-1241/Region: nucleotide-binding motif A (P-loop)
F:1316-1321/Region: nucleotide-binding motif B
F:1320-1323/Region: DEXH motif
F:1620-1866/Product: nonstructural protein NS4a #status predicted <N4A>
F:1867-2017/Product: nonstructural protein NS4b #status predicted <N4B>
F:2018-3033/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,233,299,305,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,23

Query Match 11.1%; Score 201.5; DB 1; Length 3033;
Best Local Similarity 46.2%; Pred. No. 3.9e-05;
atches 60; Conservative 13; Mismatches 40; Indels 17; Gaps 5;

QY 229 KPALVDKELVYQYDEMEAFASRGHNAFTHYVESDSAS--ASLPYDETRAIAGQF 285
DB 1619 KPTLTGPTPLLYR-----LGAVTNEVTLTHPVKRYATCMQADLEIMTSSWVLGCV 1670

QY 286 KEVLPASRGHNV--PRHYPESEPOVVTPDKETLVEAFEMEE-ASKALIEGOR 342
DB 1671 LAVAAYCATGCTGCTIIGRLHND--RVVAPDKETLVEAFEMEECSKALIEGOR 1727

QY 343 MAEMLSKIQ 352
DB 1728 MAEMLSKIQ 1737

RESULT 13
JC5620
Genome polyprotein - hepatitis C virus (isolate EUH1480)
N:Contains: capsid protein C; envelope protein M; hepatitisin (EC 3.4.21.98) (nonstructu
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C>Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
A:Accession: JC5620
R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.
Biochem. Biophys. Res. Commun. 236, 44-49, 1997

A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predomi
A:Reference number: JC5620; MUID:97366593; PMID:9223423
A:Accession: JC5620
A:Molecule type: mRNA
A:Residues: 1-3014 <CHA>
A:Cross-references: GB:Y13184
A:Experimental source: genotype 5a, which predominates in South Africa
A:Note: the translation of the nucleotide sequence is not complete in this paper
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein;
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EMP>
F:192-189/Product: major envelope protein E #status predicted <MEE>
F:384-408/Region: hypervariable #status predicted
F:390-730/Product: nonstructural protein NS1 #status predicted <NS1>
F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>
F:1008-1616/Product: hepatitisin #status predicted <NS3>
F:1231-1238/Region: nucleotide-binding motif A (P-loop)
F:1313-1318/Region: nucleotide-binding motif B
F:1317-1320/Region: DEXH motif
F:1617-1863/Product: nonstructural protein NS4a #status predicted <N4A>
F:1864-2014/Product: nonstructural protein NS4b #status predicted <N4B>
F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>
F:2210-2249/Region: interferon sensitivity determining #status predicted

Query Match 10.7%; Score 195; DB 1; Length 3014;
Best Local Similarity 26.6%; Pred. No. 0.00011;
Matches 80; Conservative 40; Mismatches 127; Indels 54; Gaps 10;

QY 23 RPAVLPDEPVLYQOEDEMEAS-HLPYIEGQNLAEQFQKALAFASRGHNVSPHYVE 81
DB 1692 RAIPTRDRVLVYQOEDEMECSASLPYDEARAIAGQFEKVLIGTAGAKET----- 1746

QY 82 SDAPPAIIPDREVLHREFDEMEAFASR-GNHVSPAHYVESDSASQAAPIYEQAVIAHQ 140
DB 1747 --LKPAAAT-----SMSSKAEQFPAKHMNMNVSGIQVLAGSTLPCNPAAV----- 1788

QY 141 FEKVLAFASRGHNDSPHYVESDAPPAIIPDREVLHREFDEMEESCHLPYIEGQMLLA 200
DB 1789 --ATLMSFPAATVSTPLTTH-----QTLLFNILGQWVASQIAP-----PTAA 1827

QY 201 EQFKKALAFASRGHNAFTHYVE-----SDAPALVPDEKLYQYDEMEAFASRG 254
DB 1828 TAFVSGMAAGAVNG-IGLGRVLIDILAGYGVGVALVAFK-INCGRPAETELVNLPL 1885

QY 255 NHVAFTHYVESDSASLPYDETRAIAGQFEKVLAFASRGHNVSPHYVESEPOVV 314
DB 1886 SLICGALVVGITCAAVLRHHIGPEGAVQMMNRILAFASRGHNSPHYVETASAKV 1945

QY 315 T 315
DB 1946 T 1946

RESULT 14
PC2219
Polyprotein - hepatitis C virus (type 5a) (fragments)
N:Contains: core protein; E1 (carboxyl end); E2/NS1 (amino end); NS3 protein; NS4A
C:Species: hepatitis C virus
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 17-Nov-2000
A:Accession: PC2219
R:Stuyver, L.; Arnhem, W.V.; Wyseur, A.; Maertens, G.
Biochem. Biophys. Res. Commun. 202, 1308-1314, 1994
A:Title: Cloning and phylogenetic analysis of the core, E2, and NS3/NS4 regions of
A:Reference number: PC2219; MUID:94338342; PMID:7520237
A:Accession: PC2219
A:Molecule type: mRNA
A:Residues: 1-876 <STU>
A:Cross-references: GB:L29577; GB:L29578; GB:L29579
A:Experimental source: serum
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: glycoprotein
F:1-191/Product: core #status predicted <COB>

F:68-78/Region: variable
 F:192-247/Product: E1 (carboxyl end) #status predicted <ERE>
 F:248-411/Product: E2/NS1 (amino end) #status predicted <ENR>
 F:248-338/Region: E2
 F:339-411/Region: NS1 (amino end)
 F:412-783/Product: NS3 #status predicted <NSR>
 F:784-837/Product: NS4A #status predicted <NSA>
 F:838-876/Product: NS4B #status predicted <NSB>
 F:281,287,294,312,340/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 10.4%; Score 188.5; DB 2; Length 876;
 Best Local Similarity 59.1%; Pred. No. 5,7e-05;
 Matches 39; Conservative 4; Mismatches 4; Indels 19; Gaps 1;

OY 229 KPALVPDKVLYOQYDEMEEPASRGNHVAPTHYVVESDASLPPYMDETRAIAGQFKER 288
 DB 817 KPALIPDREALYQOFDEME-----CSASLPPYMDETRAIAGQFKER 857

OY 289 VLAFAAS 294
 858 VLGFIS 863

RESULT 15

S35629
 hypothetical protein - hepatitis C virus

C:Species: hepatitis C virus

C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 17-Nov-2000

C:Accession: S35629

R:Sarahina, T.; Sakurai, T.; Watanabe, Y.; Kashima, K.; Suzuki, T.; Chiba, J.; Kita, Y.

Nucleic Acids Res. 21, 1037, 1993

A:Title: Nucleotide sequence of the hepatitis C virus genome from a patient negative for

A:Reference number: S35629; MID:93197128; PMID:8383835

A:Accession: S35629

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: mRNA

A:Residues: 1-125 <SAR>

A:Cross-references: EMBL:DI1353; NID:G221625; PIDN:BA01956.1; PID:G221626

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1992

C:Superfamily: hepatitis C virus genome polypeptide

Query Match 10.3%; Score 188; DB 2; Length 125;

Best Local Similarity 50.5%; Pred. No. 5e-06;

Matches 49; Conservative 10; Mismatches 32; Indels 6; Gaps 3;

OY 259 PTHYVVESDASLPPYMDETRAIAGQFKERVLAFAASRGNHVS--PRHYVESEPPQVVTVP 316

DB 2 PTKXYIATCMQADLEWMTSTRVLAGVLAVAAYCLATGCVSIIGRIHINO--RAVVAP 58

Q. 317 DKELVEAPDEME-ASKAALIEGORMAEMLSKIO 352

DB 59 DKELVEAPDEMECASKATLIEGORIAEMLSKIO 95

Search completed: July 25, 2003, 13:12:47
 Job time : 43 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 13:12:06 / Search time 52 Seconds

(without alignments)
803.912 Million cell updates/sec

Title: US-09-491-146a-52
Perfect score: 1818

Sequence: 1 AAH1PYLEQGMHLAEQFKOK.....KALIEEGQMAEWLKSQIQ 352

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 11875970 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES:

Result No.	Score	Query Match	Length	ID	Description
1	512.5	28.2	829	10 US-09-881-239-5	Sequence 5, Appl1
2	512.5	28.2	1099	10 US-09-881-654-4	Sequence 4, Appl1
3	256	14.1	2201	14 US-10-085-476-2	Sequence 2, Appl1
4	255	14.0	2985	15 US-10-259-275-40	Sequence 40, Appl1
5	255	14.0	3011	9 US-09-742-659-4	Sequence 4, Appl1
6	255	14.0	3011	11 US-09-891-894-3	Sequence 3, Appl1
7	255	14.0	3012	10 US-09-218-076-2	Sequence 2, Appl1
8	255	14.0	3012	11 US-09-995-937-2	Sequence 2, Appl1
9	255	14.0	3012	11 US-09-917-563-2	Sequence 2, Appl1
10	254	14.0	2894	10 US-09-941-611-23	Sequence 23, Appl1
11	254	14.0	2894	15 US-10-044-995-23	Sequence 23, Appl1
12	254	14.0	3011	9 US-09-916-359-2	Sequence 2, Appl1
13	251	13.8	1985	15 US-10-259-275-42	Sequence 42, Appl1
14	251	13.8	3011	16 US-10-232-643-6	Sequence 6, Appl1
15	250	13.8	3011	10 US-09-238-076-20	Sequence 20, Appl1

16	250	13.8	3011	10 US-09-952-572-9	Sequence 9, Appl1
17	250	13.8	3011	10 US-09-747-419-20	Sequence 20, Appl1
18	250	13.8	3011	11 US-09-995-937-20	Sequence 20, Appl1
19	250	13.8	3011	11 US-09-917-563-20	Sequence 20, Appl1
20	250	13.8	3011	15 US-10-259-275-20	Sequence 20, Appl1
21	249	13.7	2201	14 US-10-029-907-3	Sequence 3, Appl1
22	247	13.6	3011	10 US-09-929-955-1	Sequence 1, Appl1
23	247	13.6	3011	14 US-10-104-966-1	Sequence 1, Appl1
24	205	11.3	79	9 US-09-758-308-3	Sequence 3, Appl1
25	200	11.0	1692	11 US-09-919-901-4	Sequence 4, Appl1
26	200	11.0	1692	11 US-09-919-901-11	Sequence 11, Appl1
27	200	11.0	1692	11 US-09-919-901-18	Sequence 18, Appl1
28	200	11.0	2307	11 US-09-919-901-2	Sequence 2, Appl1
29	200	11.0	2307	11 US-09-919-901-9	Sequence 9, Appl1
30	200	11.0	2307	11 US-09-919-901-16	Sequence 16, Appl1
31	197.5	10.9	54	10 US-09-821-397-8	Sequence 8, Appl1
32	192.5	10.6	484	11 US-09-899-046-198	Sequence 198, Appl1
33	192.5	10.6	484	11 US-09-878-281-198	Sequence 198, Appl1
34	188.5	10.4	481	11 US-09-899-046-270	Sequence 270, Appl1
35	188.5	10.4	481	11 US-09-878-281-270	Sequence 270, Appl1
36	188.5	10.4	484	11 US-09-899-046-200	Sequence 200, Appl1
37	188.5	10.4	484	11 US-09-878-281-200	Sequence 200, Appl1
38	182.5	10.0	128	11 US-09-899-046-62	Sequence 62, Appl1
39	182.5	10.0	128	11 US-09-878-281-62	Sequence 62, Appl1
40	179.5	9.9	128	11 US-09-899-046-60	Sequence 60, Appl1
41	179.5	9.9	128	11 US-09-878-281-60	Sequence 60, Appl1
42	176.5	9.7	133	11 US-09-899-046-38	Sequence 38, Appl1
43	176.5	9.7	133	11 US-09-899-046-40	Sequence 40, Appl1
44	176.5	9.7	133	11 US-09-878-281-38	Sequence 38, Appl1
45	176.5	9.7	133	11 US-09-878-281-40	Sequence 40, Appl1

ALIGNMENTS

RESULT 1	US-09-881-239-5
1	Sequence 5, Application US/09881239
2	Publication No. US20020192639A1
GENERAL INFORMATION:	
1	APPLICANT: CHIEN, David Y.
2	APPLICANT: ARCANSELE, Phillip
3	APPLICANT: TANDESKE, Laura
4	APPLICANT: GEORGE-NASCIMENTO, Carlos
5	APPLICANT: COIT, Doris
6	APPLICANT: MEDINA-SELBY, Angelica
7	TITLE OF INVENTION: HCV ANTIGEN/ANTIBODY COMBINATION ASSAY
8	FILE REFERENCE: 2302-16073 / P16073.003
9	CURRENT APPLICATION NUMBER: US/09/881,239
10	CURRENT FILING DATE: 2001-06-14
11	NUMBER OF SEQ ID NOS: 8
12	SOFTWARE: PatentIn Ver. 2.0
13	SEQ ID NO 5
14	LENGTH: 829
15	TYPE: PRT
16	ORGANISM: Artificial Sequence
17	FEATURE:
18	OTHER INFORMATION: Description of Artificial Sequence: MEFA 12
19	US-09-881-239-5
Query Match	28.2% Score 512.5; DB 10; Length 829;
Best Local Similarity	Pred. No. 8.3e-35;
Matches 118; Conservative 6; Mismatches 15; Indels 49; Gaps 6;	
QY	167 KPAIIPREVLVEPEFEMEE-SQHPYIEQGMHLAEQFKOKALAFASRGHVAPTHVTE 225
DB	397 KPAIIPREVLVEPEFEMEE-SQHPYIEQGMHLAEQFKOKALG-L-SRG----- 444
QY	226 SDKPAIVPKEVLVYQYDEMEFAFASRGHVAPTHVVSASISPIYDETRALAGOF 285
DB	445 --CKPAIVPKEVLVYQYDEME-----CSQAPAYIEQAQVIAHOF 483
QY	286 KEVLAFAASGNHVSRRHYVPESEPOVVTVPKELIYEAFDENEE-ASKALIEEGQMA 344

Db 484 KEKVGLI-----DNDQVVVTPDKELLYEAFDEMEECASKALIEGQRNA 529

Qy 345 EMLKSKIQ 352

Db 530 EMLKSKIQ 537

RESULT 2

US-09-881-654-4
; Sequence 4, Application US/09881654
; Patent No. US2002014685A1
; GENERAL INFORMATION:
; APPLICANT: CHIEN, David Y.
; APPLICANT: ARANGEL, Phillip
; APPLICANT: TANDESKE, Laura
; APPLICANT: GEORGE-NASCIMENTO, Carlos
; APPLICANT: COLT, Doris
; APPLICANT: MEDINA-SELBY, Angelica
; TITLE OF INVENTION: IMMUNOSSAYS FOR ANTI-HCV ANTIBODIES
; FILE REFERENCE: 2302-17039 / PPI7039, 002
; CURRENT APPLICATION NUMBER: US/09/881,654
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: 60/212,082
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 60/280,811
; PRIOR FILING DATE: 2001-04-02
; PRIOR APPLICATION NUMBER: 60/280,867
; PRIOR FILING DATE: 2001-04-02
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1099
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MEPA 7.1
US-09-881-654-4

Query Match 28.2%; Score 512.5; DB 10; Length 1099;
Best Local Similarity 62.8%; Pred. No. 1.2e-34;
Matches 118; Conservative 6; Mismatches 15; Indels 49; Gaps 6;

Qy 167 KEAIIIPDEVLREFDEME--SOHLPYIEQGMMLAEQFOKALAFASRGNHVAPTHYVE 225
Db 701 KEAIIIPDEVLREFDEMECSQHPIYIEQGMMLAEQFOKALGL-SRG----- 748
Qy 226 SDAKPALVPDEKVLQOYDEMEEAFAASRGNHVAPTHYVESDASASLPYMETRAIAGOF 285
Db 749 --CKRAIVPDKVLLQOYDEME-----CSQAAPYIEQAQVIAHQF 787
Qy 286 KEKVLAFASRGNHVSPRHYVPESEPOVVVTPDKELLYEAFADEME--ASKALIEGQRNA 344
Db 788 KEKVLGLI-----DNDQVVVTPDKELLYEAFADEMEECASKALIEGQRNA 833
Qy 345 EMLKSKIQ 352
Db 834 EMLKSKIQ 841

RESULT 3

US-10-085-476-2
; Sequence 2, Application US/10085476
; Publication No. US20020164722A1
; GENERAL INFORMATION:
; APPLICANT: De Francesco, Raffaele
; APPLICANT: Tomei, Lucia
; APPLICANT: Behrens, Sven-Erik
; TITLE OF INVENTION: METHOD FOR REPRODUCING IN VITRO THE
; TITLE OF INVENTION: RNA-DEPENDENT RNA POLYMERASE AND TERMINAL NUCLEOTIDYL
; FILE REFERENCE: IT0002PCA
; CURRENT APPLICATION NUMBER: US/10/085,476

; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: 08/952,981

; PRIOR FILING DATE: 1998-03-23

; PRIOR APPLICATION NUMBER: PCT/IT96/00106

; PRIOR FILING DATE: 1996-05-24

; PRIOR APPLICATION NUMBER: RM95A000343

; PRIOR FILING DATE: 1995-05-25

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: fastseq for windows version 4.0

; SEQ ID NO 2

; LENGTH: 2201

; TYPE: PRT

; ORGANISM: CDNA clone pCD (38-9.4)

US-10-085-476-2

Query Match 14.1%; Score 256; DB 14; Length 2201;

Best Local Similarity 31.1%; Pred. No. 1.1e-12;

Matches 83; Conservative 31; Mismatches 75; Indels 76; Gaps 10;

Qy 23 RPAVLPDEVLQOYDEME--ASHPIYIEQGMMLAEQFOKALAF--ASRGNHVSPHY 78
Db 882 RPAVLPDEVLQOYDEMEECASHPIYIEQGMMLAEQFOKALGLQTRATQAEA----- 936
Qy 79 VPESDARPAIIPDEVLREFDEMEEAFAASR-GNHVSPHYVPESDASQAAPYIEQ----- 133
Db 937 ----AAP-----VESKRALETFAKMMNFISIQYLAGLSTLPGNPALASLMAF 984
Qy 134 -AQVIAHQFKEKVLAF-----ASRGNHVSPHYVPESDAKPAI----- 170
Db 985 TASITSPITQSTLFLNLGAVAAQLAPSAASAFVAGAGAAVSGIGKVLVDILA 1044
Qy 171 -----IPDEVLREFDEME--PYIEQGMML-----AE 201
Db 1045 GYGAGVAGALVAFKMGSEMPSTEDLVNLPALISPGALVGVCAILRRHVPGEGAV 1104
Qy 202 QFOKALAFASRGNHVAPTHYVETSDA 228
Db 1105 QMNRRLIAPASRGNHVSPHYVPESDA 1131

RESULT 4

US-10-259-275-40
; Sequence 40, Application US/10259275
; Publication No. US20030125541A1
; GENERAL INFORMATION:
; APPLICANT: Lemon, Stanley M.
; APPLICANT: Yi, Minkyung
; TITLE OF INVENTION: REPLICATION COMPETENT HEPATITIS C VIRUS AND METHODS OF USE
; FILE REFERENCE: 265,0007,0120
; CURRENT APPLICATION NUMBER: US/10/259,275
; PRIOR FILING DATE: 2003-01-13
; PRIOR APPLICATION NUMBER: US 60/171,909
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: US 09/747,419
; PRIOR FILING DATE: 2000-12-23
; PRIOR APPLICATION NUMBER: US 60/325,236
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/338,123
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: Patent In version 3.0
; SEQ ID NO 40
; LENGTH: 2985
; TYPE: PRT
; ORGANISM: ARTIFICIAL
; FEATURE:
; OTHER INFORMATION: amino acid sequence encoded by nucleotides 2077-11121 of SEQ
US-10-259-275-40

Query Match 14.0%; Score 255; DB 15; Length 2985;
Best Local Similarity 31.5%; Pred. No. 2.1e-12;
Matches 84; Conservative 31; Mismatches 74; Indels 78; Gaps 10;

RESULT 10
US-09-941-611-23
Sequence 23, Application US/09941611
Patent No. US20020106640A1
GENERAL INFORMATION:
APPLICANT: DELEYS, ROBERT J
POLLET, DIRK
MAERTENS, GEERT
VAN HEUVERSWIN, HUGO
TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
ANTIBODIES TO HEPATITIS C VIRUS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHAYE P.C.
STREET: 1100 NORTH GLEEBO ROAD
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/941,611
FILING DATE: 30-Aug-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/391,671
FILING DATE: 1995-02-21
APPLICATION NUMBER: WO PCT/EP91/02409
FILING DATE: 13-DEC-1991

RESULT 11
 US-10-044-995-23
 Sequence 23, Application US/10044995
 Publication No. US20030049665A1
 GENERAL INFORMATION:
 APPLICANT: DELEYS, ROBERT J
 POLLET, DIRK
 MAERTENS, GEERT
 VAN HEUVERSMD, HUGO
 TITLE OF INVENTION: SYNTHETIC ANTIGENS FOR THE DETECTION OF
 ANTIBODIES TO HEPATITIS C VIRUS
 NUMBER OF SEQUENCES: 23
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: NIXON & VANDERHIVE P.C.
 STREET: 1100 NORTH GLEBE ROAD
 CITY: ARLINGTON
 STATE: VA
 COUNTRY: USA
 ZIP: 22201
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/044,995
 FILING DATE: 15-Jan-2002

```
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 08/391,671
  FILING DATE: <Unknown>
  APPLICATION NUMBER: US 07/920,286
  FILING DATE: 14-OCT-1992
  APPLICATION NUMBER: WO PCT/EP91/02409
  FILING DATE: 13-DEC-1991
  APPLICATION NUMBER: EP 90124241.2
  FILING DATE: 14-DEC-1990
  ATTORNEY/AGENT INFORMATION:
    NAME: SADOFF, B.J.
    REGISTRATION NUMBER: 36,663
    REFERENCE/DOCKET NUMBER: 1487-5
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 7038164000
    TELEFAX: 7038164100
  INFORMATION FOR SEQ ID NO: 23:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 2894 amino acids
      TYPE: amino acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: peptide
      HYPOTHEICAL: NO
      ANTI-SENSE: NO
    SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-10-044-995-23

Query Match      14.0% Score 254; DB 15; Length 2894;
Best Local Similarity 31.8%; Pred. No. 2,4e-12;
Matches 85; Conservative 27; Mismatches 77; Indels 78; Gaps 10;

QY 23 RPAVIPPDEVLVYOEFDMEEAS-HLPIYEOGMQLAEQFKOKALAF---ASRGNHVSPTHY 78
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1691 KPAIIPDEVLVYREFDEMECSQHLPIYEOGMQLAEQFKOKALGLLOTASRQAEV----- 1745
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 79 VPESDARPAIIPDEVLVYREFDEMEEAFASR-GNHVSPAHYVPESDASQAAPYIEQ---- 133
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1746 ----IAPAV-----QTMQKLETFWAGHMNMFISGIQYLAGLSTLPENPAIASLMAF 1793
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 134 -AOVIAHQFKEKVLAF-----ASRGNHDSPTHYVPESDARPAI----- 170
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1794 TAAVTSPTTSQTLLFNILGQWVAQAQLAAPGAATAFVAGLAGAGVGLGVLDILA 1853
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 171 -----IPDEVLVYREFDEMEESQHL--PYIEQGMVL-----AE 201
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1854 GYGAGVAGALVAFKIMSGEVPSTEDLVNLLPALISGALVGVCAALLRRHVGPGGAV 1913
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 202 QFKOKALAFASRGNHVAPTHYVYESDA 228
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1914 QMNRRLIAFASRGNHVSPTHYVYESDA 1940
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 12
US-09-916-359-2
  Sequence 2, Application US/09916359
  Patent No. US20020034734A1
  GENERAL INFORMATION:
    APPLICANT: Veronique Barban
    TITLE OF INVENTION: VACCINE COMPOSITION FOR PREVENTING OR
    TREATING C HEPATITIS
    FILE REFERENCE: PWC97-03A
    CURRENT APPLICATION NUMBER: US/09/916,359
    CURRENT FILING DATE: 2001-07-26
    PRIOR APPLICATION NUMBER: 09/388,874
    PRIOR FILING DATE: 1999-09-02
    PRIOR APPLICATION NUMBER: 97/02,887
    PRIOR FILING DATE: 1997-03-06
    NUMBER OF SEQ ID NOS: 2
    SOFTWARE: FastSeq for Windows Version 3.0
  SEQ ID NO 2
  LENGTH: 3011
```

```
TYPE: PRT
ORGANISM: Virus
US-09-916-359-2

Query Match      14.0% Score 254; DB 9; Length 3011;
Best Local Similarity 31.8%; Pred. No. 2,4e-12;
Matches 85; Conservative 27; Mismatches 77; Indels 78; Gaps 10;

QY 23 RPAVIPPDEVLVYOEFDMEEAS-HLPIYEOGMQLAEQFKOKALAF---ASRGNHVSPTHY 78
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1691 KPAIIPDEVLVYREFDEMECSQHLPIYEOGMQLAEQFKOKALGLLOTASRQAEV----- 1745
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 79 VPESDARPAIIPDEVLVYREFDEMEEAFASR-GNHVSPAHYVPESDASQAAPYIEQ---- 133
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1746 ----IAPAV-----QTMQKLETFWAGHMNMFISGIQYLAGLSTLPENPAIASLMAF 1793
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 134 -AOVIAHQFKEKVLAF-----ASRGNHDSPTHYVPESDARPAI----- 170
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1794 TAAVTSPTTSQTLLFNILGQWVAQAQLAAPGAATAFVAGLAGAGVGLGVLDILA 1853
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 171 -----IPDEVLVYREFDEMEESQHL--PYIEQGMVL-----AE 201
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1854 GYGAGVAGALVAFKIMSGEVPSTEDLVNLLPALISGALVGVCAALLRRHVGPGGAV 1913
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 202 QFKOKALAFASRGNHVAPTHYVYESDA 228
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1914 QMNRRLIAFASRGNHVSPTHYVYESDA 1940
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 13
US-10-259-275-42
  Sequence 42, Application US/10259275
  Publication No. US20030125541A1
  GENERAL INFORMATION:
    APPLICANT: Lemon, Stanley M.
    TITLE OF INVENTION: REPLICATION COMPETENT HEPATITIS C VIRUS AND METHODS OF USE
    FILE REFERENCE: 265,0007 0120
    CURRENT APPLICATION NUMBER: US/10/259,275
    CURRENT FILING DATE: 2003-01-13
    PRIOR APPLICATION NUMBER: US 60/171,909
    PRIOR FILING DATE: 1999-12-23
    PRIOR APPLICATION NUMBER: US 09/747,419
    PRIOR FILING DATE: 2000-12-23
    PRIOR APPLICATION NUMBER: US 60/325,236
    PRIOR FILING DATE: 2001-09-27
    PRIOR APPLICATION NUMBER: US 60/338,123
    PRIOR FILING DATE: 2001-11-13
    NUMBER OF SEQ ID NOS: 73
    SOFTWARE: PatentIn version 3.0
  SEQ ID NO 42
  LENGTH: 1985
  TYPE: PRT
  ORGANISM: ARTIFICIAL
  FEATURE:
    OTHER INFORMATION: amino acid sequence encoded by the nucleotides 2119-8073 of
    US-10-259-275-42
```

```
Query Match      13.8% Score 251; DB 15; Length 1985;
Best Local Similarity 31.0%; Pred. No. 2,6e-12;
Matches 83; Conservative 30; Mismatches 75; Indels 80; Gaps 10;

QY 23 RPAVIPPDEVLVYOEFDMEEAS-HLPIYEOGMQLAEQFKOKALAF---ASRGNHVSPTHY 78
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 666 KPAIIPDEVLVYREFDEMECSQHLPIYEOGMQLAEQFKOKALGLLOTATQAE----- 720
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 79 VPESDARPAIIPDEVLVYREFDEMEEAFASR-GNHVSPAHYVPESDASQAAPYIEQ--- 133
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 721 ----AAVVESKMTL-----EAFWAGHMNMFISGIQYLAGLSTLPENPAIASLMA 767
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 134 -AOVIAHQFKEKVLAF-----ASRGNHDSPTHYVPESDARPAI----- 170
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```


Db 768 FTASITSPITTOHTLLFNILGWAQAPPSAASFVAGIAGAAVSGIGLKVLDIL 827
Qy 171 -----IPDREVLREPFDEMEBSOHL--PYIQGKML-----A 200
Db 828 AGYGAGVAGALVAFKVMSEMPSTEDLVLLPAILISPGALVGVCAILRRHVGEGCA 887
Qy 201 EQFKOKALAFASRGHVAPTHVYTESDA 228
Db 888 VQMMNRLLIAFASRGHVSPTHVYTESDA 915

RESULT 14
US-10-232-643-6
; Sequence 6, Application US/10232643
; Publication No. US20030129586A1
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, MICHAEL
; CHOO, QUI-LIM
; HAN, JANG
; CHO, JOONHO
; TITLE OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING
; HELICASE ACTIVITY AND IMPROVED SOLUBILITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHIRON CORPORATION
; STREET: Intellectual Property - R440, P.O. Box 8097
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/232,643
; FILING DATE: 30-Aug-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/483,799
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/529,169
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-3274
; TELEFAX: (510) 655-3542
; TELEX: N/A
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3011 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 9
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Lys or Arg"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 11
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Asn or Thr"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 176
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Ile or Thr"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 334
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Met or Val"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 603
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Leu or Ile"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 848
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Tyr or Asn"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 1114
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Pro or Ser"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 1117
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Ser or Thr"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 1276
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Pro or Leu"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 1454
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Cys or Tyr"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 1471
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Thr or Ser"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 1877
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Glu or Gly"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 1948
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Leu or His"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 1949
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Ser or Cys"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2021
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Gly or Val"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2349
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Thr or Ser"
; FEATURE:
; NAME/KEY: Duplication
; LOCATION: 2385
; OTHER INFORMATION: /note= "There exists a
; heterogeneity at this position - Xaa = Tyr or Phe"
; FEATURE:
; NAME/KEY: Duplication

```

1 APPLICATION NUMBER: US/09/238,076
2
3 FILING DATE:
4
5 CLASSIFICATION:
6
7 PRIOR APPLICATION DATA:
8
9 APPLICATION NUMBER: US 09/034,756
10
11 FILING DATE:
12
13 ATTORNEY/AGENT INFORMATION:
14
15 NAME: HOLLAND, DONALD R.
16
17 REGISTRATION NUMBER: 35,197
18
19 REFERENCE/DOCKET NUMBER: 6029-4831
20
21 TELECOMMUNICATION INFORMATION:
22
23 TELEPHONE: 314-727-5188
24
25 TELEFAX: 314-727-6092
26
27 INFORMATION FOR SEQ ID NO: 20:
28
29 SEQUENCE CHARACTERISTICS:
30
31 LENGTH: 3011 amino acids
32
33 TYPE: amino acid
34
35 STRANDEDNESS: single
36
37 TOPOLOGY: linear
38
39 MOLECULE TYPE: protein
40
41 HYPOTHETICAL: NO
42
43 FRAGMENT TYPE: N-terminal
44
45 US-09-238-076-20

```

	Matches	85;	Conservative	26;	Mismatches	78;	Indels	78;	Gaps	10;
QY	23	RPAVIPPREVLYOEFDEMEEAS--HLPIYEOMOLAEFOKOKALAF----	ASFONHVSPTHY	78						
Db	1691	KPRLIPREVLXOEFDMEESCOTLPRIEOGMMIAEFOKOKALLOLTQTASHNAEV-----	1745							
QY	79	VPESDARPAIIPREVLRFEDEMEAFASR-GNHVPANHVPESDSQAAPYEO----	133							
Db	1746	----ITPAV-----QTWQKLVEFMAGHMNNFIIGIQYLALSTLPGRPALASLMAF	179							
QY	134	-AOVIHQFEEKULAF-----ASGNHDSPTHYVPESDAKPAI-----	170							
Db	1794	TAAVTSPLTGQLTFNIIIGGWAAOLAABGAATAFAGAGIAGAIAISVGKLVLDILA	185							
QY	171	-----IPDREVLRFEDMEESCOHL--PYIEOCMML-----	AE 201							
Db	1854	GYGAVAGALVARPKINGSEVPSTEDLVNLPLRALSPPALVVGVCSAIIRRHVGBEGAV	191							
QY	202	QFKOKALAFASRGNHVAFTHYVESDA	228							
Db	1914	QMNRRLIAFASRGNHVSPTHYVESDA	1940							

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: July 25, 2003, 13:08:21 ; Search time 30 Seconds
(without alignments)
496.447 Million cell updates/sec

Title: US-09-491-146a-52
1818

Sequence: 1 AAHIFYLEQGMHLEQFKOK.....KALIEGQMAEMLSKIQ 352

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

a1 number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	Query Match	Length	ID	Description
1	1818	100.0	352	US-08-921-887-52	Sequence 52, Appl
2	257	14.1	3011	US-08-188-281B-1	Sequence 1, Appl
3	257	14.1	3011	US-08-453-552-1	Sequence 1, Appl
4	257	14.1	3011	US-08-710-637-1	Sequence 1, Appl
5	257	14.1	3011	PCT-US93-00907-1	Sequence 1, Appl
6	257	14.1	3011	PCT-US94-07280-1	Sequence 1, Appl
7	257	14.1	3011	PCT-US95-01087-1	Sequence 1, Appl
8	256	14.1	2013	US-08-324-977-12	Sequence 12, Appl
9	256	14.1	2013	US-08-384-616-12	Sequence 12, Appl
10	256	14.1	2013	US-08-904-686A-12	Sequence 12, Appl
11	256	14.1	2013	US-09-315-850-12	Sequence 12, Appl
12	256	14.1	2201	US-08-952-981A-2	Sequence 2, Appl
13	256	14.1	2620	US-08-324-977-32	Sequence 32, Appl
14	256	14.1	2620	US-08-384-616-32	Sequence 32, Appl
15	256	14.1	2620	US-08-904-686A-32	Sequence 32, Appl
16	256	14.1	2620	US-09-315-850-32	Sequence 32, Appl
17	256	14.1	2621	US-08-324-977-36	Sequence 36, Appl
18	256	14.1	2621	US-08-384-616-36	Sequence 36, Appl
19	256	14.1	2621	US-08-904-686A-36	Sequence 36, Appl
20	256	14.1	2621	US-09-315-850-36	Sequence 36, Appl
21	256	14.1	3010	US-08-324-977-2	Sequence 2, Appl
22	256	14.1	3010	US-08-324-977-14	Sequence 2, Appl
23	256	14.1	3010	US-08-384-616-2	Sequence 2, Appl
24	256	14.1	3010	US-08-384-616-14	Sequence 2, Appl
25	256	14.1	3010	US-08-904-686A-2	Sequence 2, Appl
26	256	14.1	3010	US-08-904-686A-14	Sequence 2, Appl
27	256	14.1	3010	US-09-315-850-2	Sequence 2, Appl

28	256	14.1	3010	US-09-315-850-14	Sequence 14, Appl
29	255	14.0	3012	US-08-811-566-2	Sequence 2, Appl
30	255	14.0	3012	US-09-034-756-2	Sequence 2, Appl
31	254	14.0	360	US-08-850-328-4	Sequence 4, Appl
32	254	14.0	859	US-08-444-818-30	Sequence 30, Appl
33	254	14.0	2261	US-08-444-818-66	Sequence 66, Appl
34	254	14.0	2772	US-08-444-818-89	Sequence 89, Appl
35	254	14.0	2894	US-08-466-975A-23	Sequence 23, Appl
36	254	14.0	2894	US-08-391-671A-23	Sequence 23, Appl
37	254	14.0	2894	US-08-467-902A-23	Sequence 23, Appl
38	254	14.0	2894	US-09-275-265-23	Sequence 23, Appl
39	254	14.0	2894	US-09-941-611-23	Sequence 23, Appl
40	254	14.0	2995	US-08-444-818-138	Sequence 138, Appl
41	254	14.0	3011	US-08-440-103-36	Sequence 36, Appl
42	254	14.0	3011	US-08-440-542-36	Sequence 36, Appl
43	254	14.0	3011	US-07-910-760-10	Sequence 10, Appl
44	254	14.0	3011	US-08-440-519-10	Sequence 10, Appl
45	254	14.0	3011	US-08-231-368-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1
US-08-921-887-52
Sequence 52, Application US/08921887
Parent No. 6030771
GENERAL INFORMATION:
APPLICANT: KHUDYAKOV, YURI E.
APPLICANT: FIELDS, HOWARD A.
TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
TITLE OF INVENTION: ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING THE SAME
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: JONES & ASKEW, LLP
STREET: 191 Peachtree Street, N.W., 37th Floor
CITY: Atlanta
STATE: GA
COUNTRY: USA
ZIP: 30303-1769
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/921,887
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: WARREN, WILLIAM L.
REGISTRATION NUMBER: 36,714
REFERENCE/DOCKET NUMBER: 03063-0380
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-818-3700
TELEFAX: 404-818-3799
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 352 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: Hepatitis virus
US-08-921-887-52

Query Match 100.0%; Score 1818; DB 3; Length 352;
Best Local Similarity 100.0%; Pred. No. 7,6e-163;
Matches 352; Conservative 0; Mismatches 0; Gaps 0;

QY 1 AAHPIYIEOGMHLAEQFKOKALPRAVTPDREVLVYOEFDMEEAASHLPYIEOGMQLAEQFK 60
DB 1 AAHPIYIEOGMHLAEQFKOKALPRAVTPDREVLVYOEFDMEEAASHLPYIEOGMQLAEQFK 60
QY 61 OKALAFASRGNHVSPTHYVESDARPAIIPDREVLHREFDMEEAFAFASRGNHVSPTHYV 120
DB 61 OKALAFASRGNHVSPTHYVESDARPAIIPDREVLHREFDMEEAFAFASRGNHVSPTHYV 120
QY 121 ESASQAAPYIEQAQVIAHOFKEKVLAFASRGNHVSPTHYVESDARPAIIPDREVLVY 180
DB 121 ESASQAAPYIEQAQVIAHOFKEKVLAFASRGNHVSPTHYVESDARPAIIPDREVLVY 180
QY 181 FDEMEESQHLPIYIEOGMHLAEQFKOKALAFASRGNHVAPTHYVTESDAKALVPDKEVLY 240
DB 181 FDEMEESQHLPIYIEOGMHLAEQFKOKALAFASRGNHVAPTHYVTESDAKALVPDKEVLY 240
QY 241 QOYDEMEEAFAFASRGNHVAPTHYVESDASASLPYMDETRAIAGOFKEKVLAFASRGNHVS 300
DB 241 QOYDEMEEAFAFASRGNHVAPTHYVESDASASLPYMDETRAIAGOFKEKVLAFASRGNHVS 300
QY 301 PRTYVSESEPOVVVTDPKELVYAFDMEEAFAFASRGNHVAPTHYVTESDAKALVPDKEVLY 352
DB 301 PRTYVSESEPOVVVTDPKELVYAFDMEEAFAFASRGNHVAPTHYVTESDAKALVPDKEVLY 352

RESULT 2

US-08-188-281B-1
Sequence 1, Application US/08188281B
Patent No. 5610009
GENERAL INFORMATION:
APPLICANT: MATANABE, SHINICHI
APPLICANT: YAMAGUCHI, JULIE
APPLICANT: DESAI, SURESH M.
APPLICANT: DEVARE, SUSHIL G.
TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
TITLE OF INVENTION: ENVELOPE GENES
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
STREET: ONE ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/188,281B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: FOREMSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 5521.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-6365
TELEFAX: 708-937-6365
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3011 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-188-281B-1

Query Match 14.1%, Score 257, DB 1, Length 3011;
Best Local Similarity 32.5%, Pred. No. 7,4e-15;
Matches 87, Conservative 27, Mismatches 74, Indels 80, Gaps 10;

QY 23 RPAVTPDREVLVYOEFDMEEAASHLPYIEOGMQLAEQFKOKALAFASRGNHVSPTHY 78
DB 1691 RPAVTPDREVLVYOEFDMEEAASHLPYIEOGMQLAEQFKOKALAFASRGNHVSPTHY 1745
QY 79 VESDARPAIIPDREVLHREFDMEEAFAFASRGNHVSPTHYVESDARPAIIPDREVLVY 133
DB 1746 VESDARPAIIPDREVLHREFDMEEAFAFASRGNHVSPTHYVESDARPAIIPDREVLVY 1792
QY 134 AQAIAHOFKEKVLAFASRGNHVSPTHYVESDARPAIIPDREVLVY 170
DB 1793 AQAIAHOFKEKVLAFASRGNHVSPTHYVESDARPAIIPDREVLVY 1852
QY 171 AQAIAHOFKEKVLAFASRGNHVSPTHYVESDARPAIIPDREVLVY 200
DB 1853 AQAIAHOFKEKVLAFASRGNHVSPTHYVESDARPAIIPDREVLVY 1912
QY 201 EGFKOKALAFASRGNHVAPTHYVTESDAKALVPDKEVLY 228
DB 1913 EGFKOKALAFASRGNHVAPTHYVTESDAKALVPDKEVLY 1940

RESULT 3

US-08-453-552-1
Sequence 1, Application US/08453552
Patent No. 5667992
GENERAL INFORMATION:
APPLICANT: CASEY, JAMES M.
APPLICANT: BODE, SUZANNE L.
APPLICANT: ZECK, BILLY J.
APPLICANT: YAMAGUCHI, JULIE
APPLICANT: FRAIL, DONALD E.
APPLICANT: DESAI, SURESH M.
APPLICANT: DEVARE, SUSHIL G.
TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
TITLE OF INVENTION: PROTEINS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
STREET: ONE ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/453,552
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: FOREMSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 5131.US.D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-9556
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3011 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-453-552-1

Query Match 14.1%, Score 257, DB 1, Length 3011;
Best Local Similarity 32.5%, Pred. No. 7,4e-15;
Matches 87, Conservative 27, Mismatches 74, Indels 80, Gaps 10;

QY	23	RPAYIPDREVLVYOEEDMEEAS- HLPYIEOGMQLAEOKOKALAF	---ASRONHVSPTH	78
Db	1691	KPAIIPDREVLVYOEEDMEECSQHLPYIEOGMQLAEOKOZALGLOTTASQAEV	-----	1745
QY	79	VPESDARPAIIPDREVLHREFDEMEAFASR- GNHVS PAHYVPESDASQAAPYIEO	---	133
Db	1746	-----ITPAVOTNNOKL----- EAWWAKMNMNFIISTGYLAGLSTLPGNPAIASIMA		1792
QY	134	--AOVIAHQFEXKYLAF----- ASRGNHDSPTHYVPESDARPAI	-----	170
Db	1793	FTAAVTSLTTSQTLIPNITLGGWAAQLAAGCAATAFAGAGIACAAISGVLGKVLDTL		1855
QY	171	-----IPDREVLVREFDEMEESQHL- PYIEOGMML	-----A	200
Db	1853	AGYAGVAGALVAFAKIMSGEVPSTEDLVNLLPAILSPALVGVCAAILRRHVGEGEA		1912
QY	201	EOFKOKALAFASRGNHVAPTHYVNESDA	228	
Db	1913	VQMMNRLIAFASRGNHVSPTHYVPESDA	1940	

· SULT 4
US-08-710-637-1

GENERAL INFORMATION:

APPLICANT: CASEY, JAMES M.
APPLICANT: BODE, SUZANNE L.

APPLICANT: BODE, SUZANNE L.

APPLICANT: ZECK, BILLY J.
APPLICANT: YAMAGUCHI, JULIE

APPLICANT: YAMAGUCHI, JULIE

APPLICANT: FRAIL, DONALD E.
APPLICANT: DESAI, SURESH M.
APPLICANT: DEVARE, SUSHIL G

TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HIV
 TITLE OF INVENTION: PROTEINS
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:

ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
STREET: ONE ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: IL

COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk

```

;
; MODERN FILE COPY CLAS
;
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
;
COMMENT: ADDITIONAL DATA

```

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/710,637
 FILING DATE:
 CLASSIFICATION: 435

; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/144,099
 ; FILING DATE:
 ; APPLICATION NUMBER: US 07/830,024
 ;

FILING DATE: 01-JAN-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: POREBSKI, PRISCIL
 REGISTRATION NUMBER: 33,

REFERENCE/DOCKET NUMBER: 5131.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-6365
TELEFAX: 708-937-9556

Query Match	14.1%;	Score 237;	DB 2;	Length 3011;
Best Local Similarity	32.5%;	Pred. No. 7.4e-15;		
Matches	87;	Conservative	27;	Mismatches 74; Indels 80; Gaps 10;

QY		23	RPAVTPREVLYOEFDMEEAS--HLPTIEOGMOLAEQFKKALAF----	ASRONHVSPTH	78
Db		1691	KPALIPREVALYEFDEMECSQHLPITIEGMMLEAFKQKALGLLOTASQAELV----		1745
QY		79	VPESDARPAIIPREVLRHEFDEMEAFASR--GNHVSAPHYVPESDAQAPYIEQ--		133
Db		1746	-----IRPAVOTWQOKL-----EAFMAKHMMNFIISGTQYLGLSTLPGNPAAIASLMA		1792
QY		134	--AQVIHAHQEKVKLF-----ASRGNHDSPTHVPESDAKPAL-----		170
Db		1793	FTAAVTSPLTTSQTLLFNLTIGGVVAOLAAPGAATTAIVGAGLAGAIGSVLGKVLVDIL		1852
QY		171	-----IPREVLYREFDEMEBSQHL--PYIEGGML-----	A	200
Db		1853	AGYGAGVAGALVAFKIMSGEVPSTEDLVNLLPAILSGALVVGVCAAILRRHVGPGEA		1912
QY		201	EQFKQKALLAFASRGNHAPHYVTESDA	228.	
Db		1913	VQWMNRLLIAFSASRGNHVSPHYVPESDA	1940	

RESULT 5
PCT-US93-00907-1

```

: Sequence 1, Application PC/TUS9300907
: GENERAL INFORMATION:

```

APPLICANT: CASEY, JAMES M.
APPLICANT: BODE, SUZANNE L.

APPLICANT: BODE, SUZANNE L.

APPLICANT: ZECK, BILLY J.
APPLICANT: YAMAGUCHI, JULIE

APPLICANT: YAMAGUCHI, JULIE

;	APPLICANT:	FRAIL, DONALD E.
;	APPLICANT:	DESAI, SURESH M.
;	APPLICANT:	DEVARE, SUSHIL G.

```

; TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:

```

ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
STREET: ONE ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: IL

```

; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIM TYPE: floppy disk

```

```

? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? Create Application Data

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: . PCT/US93/00907
FILING DATE: 19930129
CLASSIFICATION:

```

```

;
; ATTORNEY/AGENT INFORMATION:
;
; NAME: POREMBSKI, PRISCILLA E
;
; REGISTRATION NUMBER: 33,207
;
; REFERENCE/DOCKET NUMBER: 51313
;

```

TELECOMMUNICATION INFORMATION
TELEPHONE: 708-937-6365
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 1:

```

: SEQUENCE CHARACTERISTICS:
: LENGTH: 3011 amino acids
: TYPE: AMINO ACID
: STRANDEDNESS: single

```

```

;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
;      PCT-US93-00907-1

Query Match      14.1%; Score 257; DB 5; Length 3011
Best Local Similarity 32.5%; Pred. No. 7.4e-15;

```

Matches 87; Conservative 27; Mismatches 74; Indels 80; Gaps 10;

QY 23 RPAVIPDREVLYOEFDEMEAS-HLPIYEQMOLAEQFKOKALAF--ASRONHVSPTHY 78
Db 1691 KPAIIPDREVLYOEFDEMECSOHLPIYEQMMLAEQFKOALGLQTSRQAEV----- 1745

QY 79 VPESDARPAIIPDREVLYHREFDEMEAFASR--GNHVSFAHYVPESDASQAAPYIEQ--- 133
Db 1746 -----ITPAVQTNMOKL-----EAFMAKHMNFISGTQYLAGLSTLPGNPAIASLMA 1792

QY 134 --AOVIAHQFEKVLAF-----ASRGHDSPTHYVPESDAKPAI----- 170
Db 1793 FTAAVTSPLTTSQTLFNILGGVAAQLAAPGAATAFVGAGLAGAIGSVGLKVLVDIL 1852

QY 171 -----IPDREVLYREFDEMECSOHL--PYIEQGMML-----A 200
Db 1853 AGYGAGVAGALVAFKIMSGEVPSTEDLVNLLPAIISPGLVGVVCAAILRRHVGPGEA 1912

QY 201 EQFKOKALAFASRGHVAAPTHYTESDA 228
Db 1913 VQMMNRLIAFASRGHVSPTHYVPESDA 1940

RESULT 6

PCT-US94-07280-1
; Sequence 1, Application PC/TUS9407280
; GENERAL INFORMATION:
; APPLICANT: YAMAGUCHI, SHINICHI
; APPLICANT: YAMAGUCHI, JULIE
; APPLICANT: DESAI, SURESH M.
; APPLICANT: DEVARE, SUSHIL G.
; TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
; NUMBER OF SEQUENCES: 22
; TITLE OF INVENTION: ENVELOPE GENES
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07280
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5521.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3011 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-07280-1

Query Match 14.1%; Score 257; DB 5; Length 3011;
Best Local Similarity 32.5%; Pred. No. 7,4e-15;
Matches 87; Conservative 27; Mismatches 74; Indels 80; Gaps 10;

QY 23 RPAVIPDREVLYOEFDEMEAS-HLPIYEQMOLAEQFKOKALAF--ASRONHVSPTHY 78
Db 1691 KPAIIPDREVLYOEFDEMECSOHLPIYEQMMLAEQFKOALGLQTSRQAEV----- 1745

QY 79 VPESDARPAIIPDREVLYHREFDEMEAFASR--GNHVSFAHYVPESDASQAAPYIEQ--- 133
Db 1746 -----ITPAVQTNMOKL-----EAFMAKHMNFISGTQYLAGLSTLPGNPAIASLMA 1792

QY 134 --AOVIAHQFEKVLAF-----ASRGHDSPTHYVPESDAKPAI----- 170
Db 1793 FTAAVTSPLTTSQTLFNILGGVAAQLAAPGAATAFVGAGLAGAIGSVGLKVLVDIL 1852

QY 171 -----IPDREVLYREFDEMECSOHL--PYIEQGMML-----A 200
Db 1853 AGYGAGVAGALVAFKIMSGEVPSTEDLVNLLPAIISPGLVGVVCAAILRRHVGPGEA 1912

QY 201 EQFKOKALAFASRGHVAAPTHYTESDA 228
Db 1913 VQMMNRLIAFASRGHVSPTHYVPESDA 1940

RESULT 7

PCT-US95-01087-1
; Sequence 1, Application PC/TUS9501087
; GENERAL INFORMATION:
; APPLICANT: YAMAGUCHI, SHINICHI
; APPLICANT: YAMAGUCHI, JULIE
; APPLICANT: DESAI, SURESH M.
; APPLICANT: DEVARE, SUSHIL G.
; TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/01087
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 5521.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3011 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-01087-1

Query Match 14.1%; Score 257; DB 5; Length 3011;
Best Local Similarity 32.5%; Pred. No. 7,4e-15;
Matches 87; Conservative 27; Mismatches 74; Indels 80; Gaps 10;

QY 23 RPAVIPDREVLYOEFDEMEAS-HLPIYEQMOLAEQFKOKALAF--ASRONHVSPTHY 78
Db 1691 KPAIIPDREVLYOEFDEMECSOHLPIYEQMMLAEQFKOALGLQTSRQAEV----- 1745

QY 79 VPESDARPAIIPDREVLYHREFDEMEAFASR--GNHVSFAHYVPESDASQAAPYIEQ--- 133
Db 1746 -----ITPAVQTNMOKL-----EAFMAKHMNFISGTQYLAGLSTLPGNPAIASLMA 1792

QY 134 --AOVIAHQFKEKVLAF-----ASRGNHSPHYVPESDAKPAI----- 170
 DB 1793 FTAATVSPITTSQTLFFNLGCVAAQLAPGATAFVAGAGLAGAAGISVGKLVLDL 1852
 QY 171 -----IPREVLYREFDEMEESQHL--PYIEQGMML-----A 200
 DB 1853 AGYAGVAGALVAFKVMGSEVPSTEDLVNLLPALISPGALVGVVCAAILRRHVGPGEA 1912
 QY 201 EOFKOKALAFASRGNHVAPTHVYTESDA 228
 DB 1913 VQMNRLIAFASRGNHVSPHYVPESDA 1940

RESULT 8
 US-08-324-977-12
 ; Sequence 12, Application US/08324977
 ; Patent No. 5747339
 ; GENERAL INFORMATION:
 ; APPLICANT: OKAYAMA, Hiroto
 ; APPLICANT: FUKE, Isao
 ; APPLICANT: MORI, Chisato
 ; APPLICANT: TAKAMIZAWA, Akahisa
 ; APPLICANT: YOSHIDA, Iwao
 ; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
 ; TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE
 ; NUMBER OF SEQUENCES: 50
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Armstrong, Westerman, Hattori, Mclelland &
 ; STREET: 1725 K St. N.W. Suite 1000
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20006
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.5 in, 1.44MB
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
 ; SOFTWARE: ASCII
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/324,977
 ; FILING DATE: 18-OCT-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 2-167466
 ; FILING DATE: 25-JUN-1990
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 2-230921
 ; FILING DATE: 31-AUG-1990
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 2-305605
 ; FILING DATE: 09-NOV-1990
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/099,706
 ; FILING DATE: 30-JUL-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/769,996
 ; FILING DATE: 02-OCT-1991
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/635,451
 ; FILING DATE: 28-DEC-1990
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Stevens-Smith, Theresa M.
 ; REGISTRATION NUMBER: 36,281
 ; REFERENCE/DOCKET NUMBER: 900703D
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 659-2930
 ; TELEFAX: (202) 887-0357
 ; TELEX: 440142
 ; INFORMATION FOR SEQ ID NO: 12:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2013 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear

MOLECULE TYPE: protein
 US-08-324-977-12
 Query Match 14.1%; Score 256; DB 1; Length 2013;
 Best Local Similarity 31.1%; Pred. No. 5.1e-15;
 Matches 83; Conservative 31; Mismatches 75; Indels 78; Gaps 10;

QY 23 RPAVDPREVLYOEEDMEE--ASHLPTYEQGMQLAEOKOKALAF---ASRGNHVSPHY 78
 DB 1691 RPAVDPREVLYOEEDMEECSHLPYIEQGMQLAEOKOKALGILQATKQAE----- 1745
 QY 79 VPESDARPAIIPREVLYREFDEMEEAFASR-GNVSPAHVVPESDAQAPYIEQ----- 133
 DB 1746 -----AAP-----VVEASKRALETFAKGMNMFISGIQLAGLSTLPGPALASLMAF 1793
 QY 134 -AOVIAHQFKEKVLAF-----ASRGNHSPHYVPESDAKPAI----- 170
 DB 1794 TASITSPITTSQTLFFNLGCVAAQLAPPSAASAFVAGAGIAGVAGISGLKLVLDILA 1853
 QY 171 -----IPREVLYREFDEMEESQHL--PYIEQGMML-----AE 201
 DB 1854 GYGAGVAGALVAFKVMGSEVPSTEDLVNLLPALISPGALVGVVCAAILRRHVGPGEAV 1913
 QY 202 QFKOKALAFASRGNHVAPTHVYTESDA 228
 DB 1914 QMNRLIAFASRGNHVSPHYVPESDA 1940

RESULT 9
 US-08-384-616-12
 ; Sequence 12, Application US/08384616
 ; Patent No. 5847101
 ; GENERAL INFORMATION:
 ; APPLICANT: OKAYAMA, Hiroto
 ; APPLICANT: FUKE, Isao
 ; APPLICANT: MORI, Chisato
 ; APPLICANT: TAKAMIZAWA, Akahisa
 ; APPLICANT: YOSHIDA, Iwao
 ; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
 ; TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE
 ; NUMBER OF SEQUENCES: 50
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Armstrong, Westerman, Hattori, Mclelland &
 ; STREET: 1725 K St. N.W. Suite 1000
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20006
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.5 in, 1.44MB
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
 ; SOFTWARE: ASCII
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/384,616
 ; FILING DATE:
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/769,996
 ; FILING DATE: 02-OCT-1991
 ; APPLICATION NUMBER: JP 2-167466
 ; FILING DATE: 25-JUN-1990
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 2-230921
 ; FILING DATE: 31-AUG-1990
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 2-305605
 ; FILING DATE: 09-NOV-1990
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/635,451
 ; FILING DATE: 28-DEC-1990
 ; ATTORNEY/AGENT INFORMATION:

Query Match	14.1%	Score 256;	DB 3;	Length 2013;
Best Local Similarity	31.1%	Pred. No. 5.1e-15;		
Matches	83;	Conservative	31;	Mismatches 75; Indels 78; Gaps 10;

QY	23	RPAVLPDREVLVYQEFDEME--ASHLPYIEQGMQLAEQFKOKALAF--ASRQNHVSPPTHY	78
Db	1691	RPAIVPPERLLYQEFDEMEBCASHLPYIEQGMQLAEQFKOKALGILQATATKQEA----	1745
QY	79	VPESDARPAIIPDEVLHREFDEMEEAASR--GNHVSFAHNVPESDASQAPRYEQ----	133
Db	1746	----AAP-----VVEASKRALETGFPAKKMMMFISGIQYLAGLSTLPGNPAIASLMFA	1793
QY	134	-ACVIAIOFKEKVLAF-----ASRGNHDSPTHYVPESDAKPAI	170
Db	1794	TASITSPLTQOSTLFLNLIGGWAAQLAPASAAAFVAGAGIAGAAGVSGIGKVLVDILA	1653
QY	171	-----IPDREVLVREFDEMEBSOHL--PYIEQGMVL-----AE	201
Db	1854	GYGAGVAGALAFPAFGMSEMPSTEDLVLLPAILSPGALVVGVCMAAILRRHNVGPGEGAV	1913
QY	202	QFKOKALAPASRGNHVAPTHYVTESDA	228

```

DB      1914 QMMNRLLAFASRGNHVSPTHYPESDA 1940

RESULT 12
US-08-952-981A-2
; Sequence 2, Application US/06952981A
; Patent No. 6183768
GENERAL INFORMATION:
APPLICANT: DE FRANCESCO, Raffaele
APPLICANT: TOMEL, Lucia
APPLICANT: BEHRENS, Sven-Erik
TITLE OF INVENTION: METHOD FOR REPRODUCING IN VITRO THE RNA-DEPENDENT RNA
TITLE OF INVENTION: POLYMERASE AND TERMINAL NUCLEOTIDYL TRANSFERASE
TITLE OF INVENTION: ACTIVITIES ENCODED BY HEPATITIS C VIRUS (HCV)
FILE REFERENCE: IT0002P
CURRENT APPLICATION NUMBER: US/08/952,981A
CURRENT FILING DATE: 1998-03-23
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 2201
TYPE: PRT
ORGANISM: CDNA clone pCD (38-9.4)
US-08-952-981A-2

Query Match      14.1%; Score 256; DB 4; Length 2201;
Best Local Similarity 31.1%; Pred. No. 5,8e-15;
Matches 83; Conservative 31; Mismatches 75; Indels 78; Gaps 10;

QY      23 RPAVTPREVLYOEFDEMEE-ASHLPYIEQGMOLAEQKOKALAF---ASRGNHVSPTHY 78
DB      882 RPAVTPRELLYQFDEMEECASHLPYIEQGMOLAEQKOKALGLTTATKQAEA----- 936
QY      79 VPESDARPAIIPREVLAHREFDEMEEAFASR-GNHVSPTHYVPESDASQAPYIEO---- 133
DB      937 ----AAP-----VVESKWRALETFAWGHMNNFISGIQYLGLSTLPNGNPALASLMAF 984
QY      134 -AQYIAQFQEKULAF-----ASGNHDSPTHYVPESDARPAI----- 170
DB      985 TASITSELTQSTLLFNILGGWAAQLAPPASAFAGACIAGAAGVSGIGLKVLDILA 1044
QY      171 -----IPDREVLVREFDEMESQHL--PYIEQGMVL-----AE 201
DB      1045 GYGAGVAGALVAFPMVSGEMPSTEDLVNLPALISPGALVGVCAAILRRHVGPEGAV 1104
QY      202 QFKOKALAFASRGNHVAPTHYTESDA 228
DB      1105 QMMNRLLAFASRGNHVSPTHYVPESDA 1131

RESULT 13
US-08-324-977-32
US-08-324-977-32
; Sequence 32, Application US/08324977
; Patent No. 5747339
GENERAL INFORMATION:
APPLICANT: OKAYAMA, Hiroto
APPLICANT: FUKU, Isao
APPLICANT: MORI, Chisao
APPLICANT: TAKAMIZAWA, Akahisa
APPLICANT: YOSHIDA, Iwao
TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESS: Armstrong, Westernman, Hattori, McLaund &
ADDRESS: Naughton
STREET: 1725 K St. N.W. Suite 1000
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 in, 1.44mb

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: July 25, 2003, 12:54:06 ; Search time 85 Seconds

(without alignments)
657.315 Million cell updates/sec

Title: US-09-491-146a-52

Perfect score: 1818
Sequence: 1 AAHPYLEQGHLEDFKQK.....KALIEGQRMWELKSIQ 352

Scoring table:

BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

al number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1818	100.0	352	20	AAV06672
2	512.5	28.2	829	23	AAE18690
3	512.5	28.2	1099	23	AAU76378
4	512.5	28.2	1099	24	ABG72262
5	268	14.7	3010	14	AAK30616
6	263	14.5	3010	16	AAK82694
7	263	14.5	3011	14	AAK34468
8	260	14.3	3010	16	AAK68864
9	259	14.2	1786	10	AAK90158

10	259	14.2	2354	14	AAK41435
11	259	14.2	2510	13	AAK29527
12	259	14.2	3011	15	AAK66995
13	258	14.2	1250	12	AAK12599
14	257	14.1	3010	15	AAK3417
15	257	14.1	3011	16	AAK86622
16	257	14.1	3011	16	AAK40119
17	257	14.1	3011	16	AAK79232
18	256	14.1	2201	16	AAK01680
19	256	14.1	3010	13	AAK20091
20	256	14.1	3010	13	AAK20111
21	256	14.1	3010	20	AAK06423
22	256	14.1	3011	16	AAK67588
23	255	14.0	1411	13	AAK29533
24	255	14.0	3011	24	AAK77397
25	255	14.0	3011	24	AAK71460
26	255	14.0	3012	23	AAK09289
27	254	14.0	858	10	AAK90146
28	254	14.0	859	10	AAK92029
29	254	14.0	1766	10	AAK2041
30	254	14.0	1771	22	AAK62631
31	254	14.0	1771	22	AAK62632
32	254	14.0	1771	22	AAK62634
33	254	14.0	1771	22	AAK62635
34	254	14.0	1892	22	AAK62636
35	254	14.0	1911	22	AAK62638
36	254	14.0	1921	22	AAK62639
37	254	14.0	1944	22	AAK62637
38	254	14.0	2201	23	ABG30583
39	254	14.0	2201	23	ABG30586
40	254	14.0	2201	23	ABG30588
41	254	14.0	2261	10	AAK90164
42	254	14.0	2301	10	AAK92047
43	254	14.0	2435	13	AAK25135
44	254	14.0	2436	10	AAK92050
45	254	14.0	2436	10	AAK92088

ALIGNMENTS

RESULT 1
AAV06672 standard; Protein; 352 AA.
AC AAV06672;
XX 17-JUN-1999 (first entry)
DT
XX Amino acid sequence of the NS4 mosaic protein.
XX Mosaic protein; antigenic peptide; nucleocapsid; NC; hepatitis; REAL;
KW restriction endonuclease assisted ligation; vaccination; NS4 antigen.
XX
XX Hepatitis C virus.
OS Synthetic.
XX
XX W09910506-A1.
XX
XX 04-MAR-1999.
XX
XX 21-AUG-1998; 98MO-US17385.
XX 25-AUG-1997; 97US-0921887.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX Fields HA, Khudiyakov YE;
XX WPI; 1999-204671/17.
XX New mosaic protein, comprising a plurality of homologous antigenic
PT peptides from different genotypes of a species - useful for

PT-NANBH virus non
HCV antigen TTN1-3
Hepatitis C virus
Portion of PT-NANBH
Blood transmissible
HCV protein cleava
HCV genomic amino
HCV sequence. Hep
HCV NS2-NS5B non-S
Non-A, non-B viral
Non-A, non-B viral
Non-A, non-B hepat
Hepatitis C virus
HCV NS4-NS5 peptid
Hepatitis C virus
Amino acid sequenc
Hepatitis C virus
ORF extending thro
Sequence encoded i
Sequence encoded i
HCV NS3s polypepti
HCV delNS3s polype
Amino acid sequenc
Amino acid sequenc
Amino acid sequenc
Amino acid sequenc
Amino acid sequenc
Amino acid sequenc
Amino acid sequenc
Hepatitis C virus
Hepatitis C virus
Hepatitis C virus
Peptide encoded by
Sequence encoded i
HCV polypeptide 1.
Sequence encoded i
Peptide encoded by

PT detecting hepatitis infection in an individual
XX
XX Example 3; Fig 18; 66pp; English.
PS
CC The invention relates to a mosaic protein, comprising a plurality of
CC homologous antigenic peptides from different genotypes of a species. The
CC antigenic peptides are from nucleocapsid (NC) proteins. A method for
CC synthesizing an artificial gene that encodes the mosaic protein is also
CC provided. The method is designated restriction endonuclease assisted
CC ligation (REAL). The mosaic protein and the artificial mosaic protein
CC are useful for detecting a hepatitis infection in an individual. The
CC mosaic gene and protein is also useful for vaccination against infection,
CC especially hepatitis C. The method of synthesizing the artificial gene
CC and the resulting mosaic protein improve the sensitivity, spectrum of
CC immunoreactivity, and antigen specificity of enzyme immunoassays. This
CC provides improved detection of hepatitis C virus. The present sequence
CC represents the amino acid sequence of the NS4 mosaic protein.
XX
SO Sequence 352 AA:

Query Match 100.0%; Score 1818; DB 20; Length 352;
Best Local Similarity 100.0%; Pred. No. 1.8e-148;
Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAHPIPLEQGMHLAEQFKOKALRPAYIPREVLYQSFDEMEASHTPYIEQGMQLAEQPK 60
DB 1 AAHPIPLEQGMHLAEQFKOKALRPAYIPREVLYQSFDEMEASHTPYIEQGMQLAEQPK 60
QY 61 QKALAFASRGNHVSPTHYVPESDARPAIIPREVLRDEMEAFASRGNHVSFAHYVP 120
DB 61 QKALAFASRGNHVSPTHYVPESDARPAIIPREVLRDEMEAFASRGNHVSFAHYVP 120
QY 121 ESDASQAPYIEQAOVIAHQFKEKVLAFASRGNHDSPTHYVPESDAKPAIIPREVLYRE 180
DB 121 ESDASQAPYIEQAOVIAHQFKEKVLAFASRGNHDSPTHYVPESDAKPAIIPREVLYRE 180
QY 181 FEMESQHLPTYIEQGMHLAEQFKOKALAFASRGNHVAPTHYVTESDAKPAIIPREVLYRE 240
DB 181 FEMESQHLPTYIEQGMHLAEQFKOKALAFASRGNHVAPTHYVTESDAKPAIIPREVLYRE 240
QY 241 QOYDEMEAFASRGNHVAPTHYVTESDASASLPYMDETRAIAGQFKEKVLAFASRGNHVS 300
DB 241 QOYDEMEAFASRGNHVAPTHYVTESDASASLPYMDETRAIAGQFKEKVLAFASRGNHVS 300
QY 301 PPHYVPESRQVVTTPDKELIYEAPDEMEASKALIEGQMAEMLSKIQ 352
DB 301 PPHYVPESRQVVTTPDKELIYEAPDEMEASKALIEGQMAEMLSKIQ 352

-SULT 2
AAE18690
ID AAE18690 standard; Protein; 829 AA.
XX
AC AAE18690;
XX
DT 17-MAY-2002 (first entry)
XX
DE Multiple epitope fusion antigen (MEFA) 12 protein.
XX
KW Hepatitis C virus; NS3/4a antigen; multiple epitope fusion antigen;
KW HCV infection; MEFA 12 protein.
XX
OS Unidentified.
XX
XX Key Location/Qualifiers
FH Misc-difference 315
FT /note= "Encoded by ATG"
FT Misc-difference 645
TT /note= "Encoded by GAG"
XX
XX WO200196875-A2.
XX
PD 20-DEC-2001.

XX
PF 14-JUN-2001; 2001WO-US19369.
XX
XX 15-JUN-2000; 2000US-212082P.
PR 02-APR-2001; 2001US-280811P.
PR 02-APR-2001; 2001US-280867P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Chien DY, Arcangel P, Tandeske L, George-Nascllemento C, Colt D;
PI Medina-Selby A;
XX
DR WPI: 2002-179522/23.
XX
XX N-PSDB: AAD29796.
PT Immunoassay solid support useful for detecting hepatitis C virus
PT infection in a biological sample, comprises at least one of HCV
PT anti-core antibody and HCV NS3/4a epitope, bound to the support
XX
PS Disclosure: Fig 7; 87pp; English.
XX
XX The present invention relates to hepatitis C virus (HCV) core antigen
CC and NS (nonstructural) 3/4a antibody combination assay that can detect
CC both HCV antigens and antibodies present in a sample using a single
CC solid matrix as well as immunoassay solid supports for use in the assay.
CC The solid support is useful for detecting HCV infection in a biological
CC sample. The present sequence is MEFA (multiple epitope fusion antigen)
CC 12 protein. This sequence is used in the exemplification of the
CC invention.
XX
SQ Sequence 829 AA:

Query Match 28.2%; Score 512.5; DB 23; Length 829;
Best Local Similarity 62.8%; Pred. No. 1.9e-35;
Matches 118; Conservative 6; Mismatches 15; Indels 49; Gaps 6;

QY 167 KPAILIPREVLYREFDEMEE-SCHLPTYIEQGMHLAEQFKOKALAFASRGNHVAPTHYVTE 225
DB 397 KPAILIPREVLYREFDEMECSQHLPTYIEQGMHLAEQFKOKALGL----- 444
QY 226 SDAPALVPDKREVLYQOYDEMEAFASRGNHVAPTHYVTESDASASLPYMDETRAIAGQF 285
DB 445 --GKPAIIPDKREVLYQOYDEME-----CSQAPYIEQAOVIAHQF 483
QY 286 KEKVLAFASRGNHVSPPHYVPESRQVVTTPDKELIYEAPDEME-ASKALIEGQMA 344
DB 484 KEKVLGLI-----DNDQVVTTPDKELIYEAPDEMECASKALIEGQMA 529
QY 345 EMLSKIQ 352
DB 530 EMLSKIQ 537

RESULT 3
AAU76378
ID AAU76378 standard; Protein; 1099 AA.
XX
AC AAU76378;
XX
DT 08-MAY-2002 (first entry)
XX
DE HCV multiple epitope fusion antigen (MEFA) 7.1 protein sequence.
XX
KW Hepatitis C virus; HCV; NS3/4a conformational epitope; seroconversion;
KW immunoassay solid support; multiple epitope fusion antigen; MEFA;
KW non-structural protein.
XX
OS Hepatitis C virus.
OS Synthetic.
XX
XX WO200196870-A2.
XX
PD 20-DEC-2001.

```

XX PF 14-JUN-2001; 2001WO-US19156.
XX PR 15-JUN-2000; 2000US-212082P.
XX PR 02-APR-2001; 2001US-280811P.
XX BR 02-APR-2001; 2001US-280867P.
XX PA (CHIR ) CHIRON CORP.
XX PI Chien DY, Arcangel P, Tandeske L, George-nascimento C, Coit D,
XX PI Medina-selby A;
XX DR WPI; 2002-090228/12.
XX DR N-PSDB; ABK15345.
XX PT Immunoassay solid support, useful for detecting hepatitis C virus
XX PT infection in biological sample, comprises HCV NS3/4a conformational
XX PT epitope and multiple epitope fusion antigen bound to the support
XX PT
XX Claim 5; Fig 5; 92pp; English.

CC The present invention relates to a new immunoassay solid support
CC consisting essentially of at least one hepatitis C virus (HCV) NS3/4a
CC conformational epitope and a multiple epitope fusion antigen (MEFA),
CC bound to the support. The NS3/4a conformational epitope and/or
CC MEFA reacts specifically with anti-HCV antibodies present in a biological
CC sample from an HCV-infected individual. The immunoassay of the invention
CC is useful for detecting hepatitis C virus infection in a biological
CC sample. The method of the invention provides a sensitive, accurate
CC diagnostic and prognostic tool to provide adequate patient care and to
CC prevent transmission of HCV by blood and by blood products, or by
CC personal contact. Use of NS3/4a conformational epitope in combination
CC with MEFA, provides a sensitive and reliable method for detecting early
CC HCV seroconversion. Use of MEFA has the added advantages of decreasing
CC masking problems, improving sensitivity in detecting antibodies by
CC allowing a greater number of epitopes on a unit surface area of
CC substrate, and improving substrate. Detection accuracy is increased and
CC the incidence of false results is reduced because of the identification
CC and the use of highly immunogenic HCV antigens which are present during
CC the early stages of HCV seroconversion. The present amino acid sequence
CC represents the multiple epitope fusion antigen (MEFA) 7.1 of the
CC invention.
CC
CC Sequence 1099 AA;
SQ
Query Match 28.2%; Score 512.5; DB 21; Length 1099;
Best Local Similarity 62.8%; Pred. No. 2,8e-35;
Matches 118; Conservative 6; Mismatches 15; Indels 49; Gaps 6;

167 KPAIIPDREVLRFDEME--SOHLPTIEQGMMLAEQFOKALAFASRGNHVAPTHVYTE 225
DB 701 KPAIIPDREVLRFDEMECSOHLPTIEQGMMLAEQFOKALGL----- 748
OY 226 SDAPALVPDEKVLVQYDDEMEAFASRGNHVAPTHVVSASLPMDETRAIAGQF 285
DB 749 --GKPAIVPDKREVLYQYDDEME-----CSQAAPVIEQAQVIAHQF 787
OY 286 KKKVLAFAFASRGNHVSPRHVPSBPQVVTTPKKEILYEFDEME--ASGALALIEGQRNA 344
DB 788 KKKVGLGLI-----DNDQVVTPDKKILYEAFAFDEMECASKALALIEGQRNA 833
OY 345 EMLKSKIQ 352
DB 834 EMLKSKIQ 841

```

```

XX DE HCV multiple epitope fusion antigen 7.1 (MEFA 7.1).
XX XX
XX XX Immunoassay solid support; Hepatitis C Virus type-1; HCV-1; HCV-2;
XX KM NS3/4a conformational epitope; multiple epitope fusion antigen 7.1;
XX KM MEFA 7.1; anti-HCV antibody; NS3/4a conformational antigen; HCV-3;
XX KM HCV infection; Hepatitis C Virus type-2; Hepatitis C Virus type-3;
XX KM mutant; mutein.
XX XX
XX OS Chimeric - Hepatitis C virus type 1.
XX OS Chimeric - Hepatitis C virus type 2.
XX OS Chimeric - Hepatitis C virus type 3.
XX OS Synthetic.
XX XX
XX Key
XX FH Location/Qualifiers
XX FT 1..156
XX FT /note= "Correspond to amino acids 1-156 of HCV-1 hSOD
XX FT superoxide dismutase"
XX FT
XX FT Region
XX FT 159..176
XX FT /note= "Correspond to amino acids 303-320 of HCV-1 E1"
XX FT 179..199
XX FT /note= "Correspond to consensus sequence of amino acids
XX FT 390-410 of HCV-1 E2 HVR"
XX FT
XX FT Region
XX FT 200..230
XX FT /note= "Correspond to consensus sequence of amino acids
XX FT 384-414 of HCV-1 and HCV-2 E2 HVR"
XX FT
XX FT Region
XX FT 231..696
XX FT /note= "Correspond to amino acids 1193-1658 of HCV-1
XX FT helicase"
XX FT 699..745
XX FT /note= "Correspond to amino acids 1689-1735 of HCV-1
XX FT 5-1-1 epitope"
XX FT 748..794
XX FT /note= "Correspond to amino acids 1689-1735 of HCV-3
XX FT 5-1-1 epitope"
XX FT 797..843
XX FT /note= "Correspond to amino acids 1689-1735 of HCV-2
XX FT 5-1-1 epitope"
XX FT 846..881
XX FT /note= "Correspond to amino acids 1901-1936 of HCV-1
XX FT polypeptide C100"
XX FT
XX FT Region
XX FT 884..919
XX FT /note= "Correspond to amino acids 2278-2313 of HCV-1
XX FT NS5 region"
XX FT 922..957
XX FT /note= "Correspond to amino acids 2278-2313 of HCV-1
XX FT NS5 region"
XX FT 958..1028
XX FT /note= "Correspond to amino acids 2278-2313 of HCV-1
XX FT NS5 region"
XX FT
XX FT Region
XX FT 1029..1099
XX FT /note= "Correspond to core region antigenic determinants
XX FT from amino acids 9-32, 39-42 and 64-88 of HCV-1
XX FT and amino acids 67-84 of HCV-2"
XX FT
XX PN US2002146685-A1.
XX XX
XX PD 10-OCT-2002.
XX XX
XX PF 14-JUN-2001; 2001US-0881654.
XX XX
XX PR 15-JUN-2000; 2000US-212082P.
XX PR 02-APR-2001; 2001US-280811P.
XX PR 02-APR-2001; 2001US-280867P.
XX XX
XX PA (CHIE//) CHIEN D Y.
XX PA (ARCA//) ARCANGEL P.
XX PA (TAND//) TANDESKE L.
XX PA (GEOR//) GEORGE-NASCIMENTO C.
XX PA (COIT//) COIT D.
XX PA (MEDI//) MEDINA-SELBY A.
XX XX

```

PI Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;
PI Medina-Selby A;
XX
XX
DR NPI; 2003-147573/14.
DR N-PSDB; ABX14411.
XX
XX
PT Immunoassay solid support for detecting Hepatitis C Virus infection in
PT biological samples, comprises Hepatitis C Virus conformational epitope
PT and multiple epitope fusion antigen -
XX
XX
PS Claim 25; Fig 5A-5F; 45pp; English.
XX
XX
CC The present invention relates to immunoassays comprising Hepatitis C
CC virus (HCV) NS3/4a conformational epitope and multiple epitope fusion
CC antigen (MEFA), bound to a solid support. The NS3/4a epitope and/or
CC the multiple epitope fusion antigen react with anti-HCV antibodies
CC present in a biological sample from an HCV-infected individual. The
CC immunoassays and methods of the invention are useful for detecting
CC HCV infection in a biological sample. The inventive immunoassay solid
CC support provides a sensitive and reliable method for detecting early
CC HCV seroconversion. The assays can detect HCV infection caused by any
CC six known genotypes of HCV. The use of the multiple epitope fusion
CC proteins decreases masking problems, improves sensitivity in detecting
CC antibodies by allowing a greater number of epitopes on a unit area
CC of substrate, and improves selectivity. The present sequence
CC represents HCV multiple epitope fusion antigen 7.1 (MEFA 7.1), a
CC mutant HCV polypeptide derived from various regions of HCV type 1,
CC 2, or 3 (HCV-1, HCV-2, or HCV-3) polypeptide sequences.
XX
XX
SQ Sequence 1099 AA;
XX
XX
Query Match 28.2%; Score 512.5; DB 24; Length 1099;
Best Local Similarity 62.8%; Pred. No. 2,8e-35;
Matches 118; Conservative 6; Mismatches 15; Indels 49; Gaps 6;
XX
QY 167 KPAIIPDREVLVYREFDEME--SOLPYIEGMMALAEQFKALAFASRGNHVAPTHYTE 225
DB 701 KPAIIPDREVLVYREFDEMECSOHLPIYIEGMMALAEQFKALGL-SRG----- 748
QY 226 SDAKPAIVDPKEVLYOQYDEMEPAFASRGNHVAPTHYVESDASLPYMDETRAIAGQF 285
DB 749 --GKPAIVDPKEVLYOQYDEME-----CSQAAPYIEQAQVIAHQF 787
QY 286 KEKVLAFASRGNHVSRHYVPESEPOVVTTPDKEILYEAFDEME--ASKALIEEQRMA 344
DB 788 KEKVLGLI-----DNDQVVTTPDKEILYEAFDEMECASRAALIEEQRMA 833
345 EMLKSKIQ 352
834 EMLKSKIQ 841
XX
XX
RESULT 5
ID AAR30616 standard; Protein; 3010 AA.
XX
XX
AC AAR30616;
XX
XX
DT 25-MAR-2003 (updated)
DT 19-MAR-1993 (first entry)
XX
XX
DE Polypeptide coded by Korean HCV full cDNA sequence LBCL.
XX
XX
KM KHCV-LBCL; diagnosis; vaccine.
XX
XX
OS Korean hepatitis C virus.
XX
XX
PN EP52J318-A2.
XX
XX
PD 07-JUN-1993.
XX
XX
PF 10-JUN-1992; 92EP-0109753.
XX

PR 10-JUN-1991; 91KR-0009510.
PR 06-AUG-1991; 91KR-0013601.
XX
XX
PA (LUCK-) LUCKY LTD.
XX
XX
PI Cho JM, Choi DY, Kim CH, Kim ST, Lee YB, Lim KJ, Park YM;
PI So HS, Yang JY;
XX
XX
DR NPI; 1993-001883/01.
DR N-PSDB; AAQ33282.
XX
XX
PT DNA and polypeptide(s) from a new type of hepatitis C virus (KHCV)
PT for diagnosing and vaccinating against KHCV infections
XX
XX
PS Disclosure; Fig 2; 11pp; English.
XX
XX
CC The polypeptide is that encoded by the full cDNA sequence of Korean
CC hepatitis C virus (KHCV) cDNA, KHCV-LBCL. It or its fragments may be
CC used in a specific and accurate method for detecting KHCV antibodies
CC in the serum of hepatitis C patients. Antibodies directed against these
CC polypeptides are useful for the purification of KHCV antigens and for
CC the development of an improved diagnostic to detect KHCV antigens in a
CC sample. The polypeptides may also be used in a vaccine for treatment
CC and prevention of KHCV infection at a dosage of 5-200 ug/peptide.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
XX
SQ Sequence 3010 AA;
XX
XX
Query Match 14.7%; Score 268; DB 14; Length 3010;
Best Local Similarity 30.7%; Pred. No. 1.3e-13;
Matches 102; Conservative 26; Mismatches 88; Indels 116; Gaps 14;
XX
QY 23 RPAVIPDREVLVYREFDEME--ASHLPYIEGMMALAEQFKALAF--ASRGNHVSPHY 78
DB 1691 KPAIIPDREVLVYREFDEMECSOHLPIYIEGMMALAEQFKALGLQTRQAQA----- 1745
QY 79 VPEASAPAIIPDREVLVYREFDEMEAFASR-GNHVSAHYVESDASQAAPYIEQAQVI 137
DB 1746 ---AAP-----VVESKMALETTPAKMMNMTISGIQVLGLSTLPNPAIRSP--- 1790
QY 138 AHQFEKVLAFASRGNHDSPTHYVESDAKPAIIPDREVLVYREFDEMEESOH-LPYIEG 196
DB 1791 -----MAFTA-----STSP-----TTQHTLLFNILG 1813
QY 197 MMLAEQFKALAFASRGNHVAPTHYTESDAK-----DALVPDKEVLYOQ 242
DB 1814 GVVAAQLAPSAASFVAGIAGAAGVITGLGKVLVDILAGYAGACVAGALVAFK-IMSGE 1872
QY 243 YDEME-----AFPSRG-----NHVAPTHYVESDASASLPYMDETRAIAG 283
DB 1873 MESAEDVNLPLPAILSPALVGVICAILRRHVGPGGAV----- 1913
QY 284 QFEKVLAFASRGNHVSPRHYVPESEPOVVT 315
DB 1914 QMMNRLLAFASRGNHVSPRHYVPESEPARPT 1945
XX
XX
RESULT 6
ID AAR82694 standard; Protein; 3010 AA.
XX
XX
AC AAR82694;
XX
XX
DT 14-NOV-1996 (first entry)
XX
XX
DE Partial HCV non-structural polypeptide.
XX
XX
KM proteinase; hepatitis C virus; screening; inhibitor; proteolytic;
KM identification; cleavage.
XX
XX
OS Hepatitis C virus.
XX
XX
FH Key Location/Qualifiers

```

FT      Protein      898..1233
FT      /note="partial proteinase; see AAR82692"
FT      Protein      992..1907
FT      /note="partial proteinase; see AAR82693"
XX      JP07184648-A.
XX      25-JUL-1995.
XX      PD
XX      PF
XX      05-FEB-1993; 93JP-0018854.
XX      PR
XX      04-DEC-1992; 92JP-0325303.
XX      07-FEB-1992; 92JP-0022657.
XX      18-SEP-1992; 92JP-0249240.
XX      (KAEN/) KAENNO K.
XX      (SOYA-) SOYAKU GIJUNSU KENKYUSHO KK.
XX      (SUMO) SUMITOMO METAL IND LTD.
XX      WPI; 1995-287962/38.
XX      N-PSDB; AAT03960.

PT      An HCV proteinase active substance - which has activity as an
PT      anti-HCV agent and can be used to screen for proteinase inhibitors
PS      Disclosure; Page 39-48; 52pp; Japanese.
XX      The present sequence is a partial Hepatitis C Virus (HCV) polyprotein
XX      from the non-structural region. Partial proteinase sequences (AAR82692-
XX      93) are contained within this sequence. The proteinases can be used as
XX      anti-HCV agents. They can also be used to screen cpds. for their ability
XX      to inhibit their proteolytic activity. In this way proteinase inhibitors
XX      can be identified.
XX      Sequence 3010 AA;

Query Match      14.5%; Score 263; DB 16; Length 3010;
Best Local Similarity 31.8%; Pred. No. 3.6e-13;
Matches 89; Conservative 24; Mismatches 63; Indels 104; Gaps 11;

QY      23 RPAVTPREVLYOEFDEMEE-ASHLPYIEQGMQLAEQKOKALAF--ASQ----- 70
DB      1691 RPAVTPREVLYOEFDEMEECASHLPYIEQGMQLAEQKOKALGLQTATQAEAAAFV 1750
QY      71 -----NHVSPTHYVPESDARPAIIPREVLYHREFDEMEAFASRGNHVS 114
DB      1751 ESKRALEVFMAKMMNFISGIQYV-----AGLSTLPGNPAT-----ASLMAFTASITS 1799
QY      115 P-----AHYVPESDAS-----QAPYIEQAVIAHQFKEKVLAF 148
DB      1800 PLTTQNTLLFNILGQVAAQLAPPSAASAFVAGAGIAGAAVSGIGKVLVH-----LLAG 1854
QY      149 ASRGNHDSPTHYVESDAKPAIIPREVLYHREFDEMESQHL--PYIEQGMML----- 199
DB      1855 YGAGVAGALVAF-----KVMSEGMSTEDLVNLLPALISFGALVGVCA 1900
QY      200 -----AEQKOKALAFASRGNHVAPTHYTESDA 228
DB      1901 ILRRHVGPGEAVQMMNRLIAFASRGNHVSPTHYVESDA 1940

RESULT 7
AAR34468
ID      AAR34468 standard; Protein; 3011 AA.
XX      AAR34468;
XX      30-JUL-1993 (first entry)
XX      Encoded by full-length Hepatitis C virus clone JKI-B.
XX      HCV; non-A, non-B hepatitis virus; NANBH; liver disease;
XX      polymerase chain reaction; diagnostic method.

```

```

XX      XX
OS      Hepatitis C virus.
XX      Key      location/Qualifiers
XX      Key      location/Qualifiers
XX      Misc-difference 2414
XX      /note="not defined"
XX      JP05068562-A.
XX      23-MAR-1993.
XX      PD
XX      PF
XX      30-MAY-1991; 91JP-0153736.
XX      PR
XX      30-MAY-1991; 91JP-0153736.
XX      (SANW) SANWA KAGAKU KENKYUSHO CO.
XX      WPI; 1993-130638/16.
XX      DR
XX      N-PSDB; AAQ04026.
XX      DNA and cDNA of hepatitis C virus - useful as probes for
XX      diagnosing HCV infection
XX      Claim 3; Page 6-18; 44pp; Japanese.
XX      cDNA was prepared from HCV genomic RNA. Full-length clone JKI-B
XX      and 14 shorter clones were isolated by PCR amplification.
XX      primer/probes derived from the sequences of these clones can be used
XX      in diagnostic assays for HCV. See also AAQ04025-Q40439.
XX      Sequence 3011 AA;

Query Match      14.5%; Score 263; DB 14; Length 3011;
Best Local Similarity 32.5%; Pred. No. 3.6e-13;
Matches 87; Conservative 27; Mismatches 74; Indels 80; Gaps 11;

QY      23 RPAVTPREVLYOEFDEMEE-ASHLPYIEQGMQLAEQKOKALAF--ASRGNHVSPTHY 78
DB      1691 RPAVTPREVLYOEFDEMEECASHLPYIEQGMQLAEQKOKALGLQTATQAEAAAFV 1745
QY      79 VPESDARPAIIPREVLYHREFDEMEEAFASR--GNHVSPTHYVPESDASQAPYIEO-- 133
DB      1746 -----AAPVSESKQAL-----EAFMAKMMNFISGIQYVAGLSTLPGNPATVSLMA 1792
QY      134 --AQVTHQKKEKVLAF-----ASRGNHDSPTHYVPESDAKPAI----- 170
DB      1793 FTASITSPLTQNTLLFNILGQVAAQLAPPSAASAFVAGAGIAGAAVSGIGKVLVDIL 1852
QY      171 -----IPREVLYREFD-EMESQHL-----PYIEQGMML-----A 200
DB      1853 AGYAGAGVAGALVAFKMGSGEMSTEDLVNLLPALISFGALVGVCAAILRRHVGPGEA 1912
QY      201 EQFKOKALAFASRGNHVAPTHYTESDA 228
DB      1913 VQMMNRLIAFASRGNHVSPTHYVESDA 1940

RESULT 8
AAR68864
ID      AAR68864 standard; Protein; 3010 AA.
XX      AAR68864;
XX      06-DEC-1995 (first entry)
XX      Hepatitis C virus RNA helicase.
XX      Hepatitis C virus; HCV; non-A non-B; helicase gene; RNA helicase;
XX      baculovirus; recombinant production.
XX      Hepatitis C virus.
XX      Key      Location/Qualifiers

```


Accession	Region	Score	DB	Length
FT	Region	196..198		
FT		/label= N-linked glycosylation site		
FT	Region	209..211		
FT		/label= N-linked glycosylation site		
FT	Region	234..236		
FT		/label= N-linked glycosylation site		
FT	Region	250..252		
FT		/label= N-linked glycosylation site		
FT	Region	305..307		
FT		/label= N-linked glycosylation site		
FT	Region	325..327		
FT		/label= N-linked glycosylation site		
FT	Region	417..419		
FT		/label= N-linked glycosylation site		
FT	Region	423..425		
FT		/label= N-linked glycosylation site		
FT	Region	430..432		
FT		/label= N-linked glycosylation site		
FT	Region	448..450		
FT		/label= N-linked glycosylation site		
FT	Region	532..534		
FT		/label= N-linked glycosylation site		
FT	Region	556..558		
FT		/label= N-linked glycosylation site		
FT	Region	576..578		
FT		/label= N-linked glycosylation site		
FT	Region	623..625		
FT		/label= N-linked glycosylation site		
FT	Region	645..647		
FT		/label= N-linked glycosylation site		
FT	Region	1213..1215		
FT		/label= N-linked glycosylation site		
FT	Region	1255..1257		
FT		/label= N-linked glycosylation site		
FT	Region	2041..2043		
FT		/label= N-linked glycosylation site		
FT	Region	2077..2079		
FT		/label= N-linked glycosylation site		
FT	Region	2240..2242		
FT		/label= N-linked glycosylation site		
FT	Region	2788..2790		
FT		/label= N-linked glycosylation site		
XX				
PN		JP06319583-A.		
XX				
PD		22-NOV-1994.		
XX				
XX		18-SEP-1992;	92JP-0249241.	
XX		18-SEP-1992;	92JP-0249241.	
XX				
PA		(SOYA-) SOYAKU GIUTSU KENKYUSHO KK.		
XX				
DR		WPI, 1995-040330/06.		
XX		N-PSDB; AA081559.		
XX				
PT		Expression of hepatitis C virus helicase gene in baculovirus -		
PT		useful for large scale prodn. of RNA helicase.		
XX				
PS		Claim 1; Fig 1-4; 9pp; Japanese.		
XX				
CC		AA081559 encodes AAR68864 hepatitis C virus (HCV) RNA helicase. The		
CC		C DNA was used in the construction of an expression vector which		
CC		was used to transform a baculovirus host. The transformed		
CC		baculovirus could then be used for the recombinant prodn. of		
CC		HCV RNA helicase.		
XX				
XX				
SO		Sequence 3010 AA;		
XX				
Query Match		14.3%; Score 260; DB 16; Length 3010;		
Best Local Similarity		32.2%; Pred. No. 6.4e-13;		
Matches		86; Conservative 28; Mismatches 75; Indels 78; Gaps 10;		

QY	23	KPAIIPREVLYOFDEMEE--ASHLPRIEOMLOAEQKALAF---ASQNHVSPHNY	78
Db	1691	RPAPIPREVLYOFDEMEECASHLPRIEOMLOAEQKALALLOTATQDAE----	1745
QY	79	VPESDARPAIIPREVLYHREFDEMEEAFASR--GNHVSRANYVPESDASQAAPYIEQ----	133
Db	1746	----AAR-----VYESKVALEVFMAKMMNFISGIQYLAGISLTPGNRALASLMAF	1799
QY	134	-AOVIAQOFKEKULAF-----ASQNHDSPIHNYVPESAKAI-----	170
Db	1794	TASITSPILTONTLFFNIIIGSWAAOQLAPSAASAFAGAGIAGAAGVSIIGKVLVDILA	1853
QY	171	-----IPREVLYHREFDEMEESQHL--PYIEQGMIL-----AE	201
Db	1854	GUGAGVAGVALFAKVMGSGEMSPSTEDLVNLLRALISPGALVUGVUCSAIIRKRVGSEGA	1913
QY	202	QFKOKALAFASRGHNHAPTHYVPSDA	228
Db	1914	QMMNRLIAFASRGHNHAPTHYVPSDA	1940

XX	RESULT 9
XX	AAP90158
ID	AAP90158 standard; protein: 1786 AA.
XX	
AC	AAP90158;
XX	
DT	25-MAR-2003 (updated)
DT	10-NOV-1989 (first entry)
XX	
DE	Protein sequence of hepatitis C virus composite cDNA.
XX	
KM	Hepatitis C virus; vaccine.
XX	
OS	Pan troglodytes.
XX	
PN	GB2212511-A.
XX	
PD	26-JUL-1989.
XX	
PF	18-NOV-1988; 88GB-0027024.
XX	
PR	18-NOV-1987; 87US-0122714.
PR	30-DEC-1987; 87US-0139886.
PR	26-FEB-1988; 88US-0161072.
PR	26-OCT-1988; 88US-0263584.
XX	
PA	(CHIR) CHIRON CORPORATION.
XX	
PI	Houghton M, Choo QL, Kuo G;
XX	
DR	WPI; 1989-215054/30.
DR	N-PSDB; AAN90327.
XX	
PT	Hepatitis C virus gene - used for prodn. of polynucleotide probes, poly(peptide(s) and antibodies for diagnosis, prevention and treatment of infection.
XX	
PS	Disclosure; fig 26; 30pp; English.
XX	
CC	The sequence is encoded by the composite cDNA of AAN90327. These
CC	antigens react with antibodies in patients with non-A non-B hepatitis
CC	(NANBH). They can be used to diagnose HCV-induced NANBH, to raise
CC	antibodies for immunoassay or treatment, or to produce vaccines.
CC	(updated on 25-MAR-2003 to correct PR field.)
XX	
SQ	Sequence 1786 AA;
XX	
Query Match	14.2%; Score 259; DB 10; Length 1786;
Best Local Similarity	31.5%; Pred. No. 3, 8e-13;
Matches 84; Conservative 28; Mismatches 77; Indels 78; Gaps 9;	
23 RPAVIDREVLYOEFDEMEERAS-HLPYIEGQMLAEOFKQKALAF--ASRONHVSPTHY 78	

CC invention. The HCV gene is useful in the development of a diagnostic
CC method which is more accurate and effective than conventional ones. In
CC the detection of antibodies raised against a wide range of HCVs which
CC have been hardly detected before. The complete gene may be used in an
CC in vitro screening system for a substance capable of specifically
CC suppressing or controlling a proteolytic processing of a precursor
CC polypeptide of HCV.
CC (Updated on 25-MAR-2003 to correct PN field.)

CC XX Sequence 2510 AA;

Query Match 14.2%; Score 259; DB 13; Length 2510;
Best Local Similarity 30.3%; Pred. No. 6, 1e-13;
Matches 96; Conservative 26; Mismatches 109; Indels 86; Gaps 11;

QY 23 RPAVIPPREVLYOEFDEMEF ASHLPIYIEQGMQLAOFKOKALAF---ASRQNHVSPTHY 78
DB 1691 RPAVIPPREVLYOEFDEMEFECASHLPYIEQGMQLAOFKOKALGLTQATKQABA----- 1745
79 VPESDARPAIIPDREVLHREFDEMEAFASR-GNHVSPAHYVPESDASQAAPYIEQOVI 137
1746 -----AAP-----VVESKRALETWAKHMMNFISGIQYLAGLSTLPQNPAL----- 1787
QY 138 AHOFKEKVLAFASRGNHDSPTHYVPESDAKPAIIPDREVLHREFDEMEESQHLPIYIEQGM 197
DB 1788 -----ASLMFTASITSPITQYT-----LNFNLGQVMAAQALAPPSASA 1828
QY 198 MLEQFQKALAFASRGNHVAPTHYTESDAKPAIIPDREVLHREFDEMEESQHLPIYIEQGM 251
DB 1829 FVAGAGITGAAGVSGISGLKVLVDILAGYAGAGALVAFK-VMSGDMSTEDLVNLLPAL 1867
QY 252 SRG-----NHVAPTHYVPESDASASLPYDETRALAGQFKEKVLAFASRGNH 298
DB 1888 SPGALVGVCAAILRRHVGPBGCAV-----GMNMLIAFASRGNH 1928
QY 299 VSPRHVYVESEPOVVVT 315
DB 1929 VSPTHVVPESDAARVT 1945

RESULT 12

AA.56995
ID AAR66995 standard; Protein; 3011 AA.
XX
AC AAR66995;
XX
DT 01-AUG-1995 (first entry)

Hepatitis C virus gene HC-J1/protein.

KM Hepatitis C virus; HCV gene HC-J1/protein; specific antibodies.

XX Hepatitis C virus.

XX JP06284887-A.

XX 11-OCT-1994.

XX 10-DEC-1993; 93JP-0345753.

XX 10-DEC-1992; 92JP-0360705.

XX (IMMO) IMMUNO JAPAN KK.

XX WPI; 1994-362594/45.

XX N-PSDB; AAQ74770.

XX HCV genes and the corresponding proteins - used in the production
PT of anti-HCV antibodies and the detection of HCV infection

XX Claim 1; Pages 18-32; 35pp; Japanese.

XX AAQ74770 encodes AAR66995 the HC-J1/protein, the cDNA can be used in
CC

CC the construction of an expression vector for the transformation
CC of a host cell. The host cell can then be used in the production
CC of proteins and peptides, useful in the preparation of monoclonal
CC and polyclonal HCV-specific antibodies.

CC XX Sequence 3011 AA;

Query Match 14.2%; Score 259; DB 15; Length 3011;
Best Local Similarity 31.7%; Pred. No. 7, 9e-13;
Matches 85; Conservative 28; Mismatches 75; Indels 80; Gaps 10;

QY 23 RPAVIPPREVLYOEFDEMEFAS-HLPYIEQGMQLAOFKOKALAF---ASRQNHV-SPTH 77
DB 1691 RPAIIPDREVLHREFDEMEESQHLPIYIEQGMQLAOFKOKALGLTQATSRQAEVIAPT- 1749
QY 78 YVESDARPAIIPDREVLHREFDEMEAFASR-GNHVSPAHYVPESDASQAAPYIEQO--- 133
DB 1750 -----VOTNQKLEAFKAKHMMNFISGIQYLAGLSTLPQNPALIASLMA 1792
QY 134 --AQVIAHOFKEKVLAF-----ASRGNHDSPTHYVPESDAKPAI----- 170
DB 1793 FTAAVTSPITTSQTLFNLGQVMAAQALAPGATAFVSGLAGAAGVSGCRVLVDIL 1852
QY 171 -----IPDREVLHREFDEMEESQHL--PYIEQGMVL-----A 200
DB 1853 AGYAGAGALVAFKIMSGELPSTEDLVNLLPALISPGALVGVCAAILRRHVGPBGCA 1912
QY 201 EGFQKALAFASRGNHVAPTHYTESDA 228
DB 1913 VQMMNMLIAFASRGNHVSPTHYVPESDA 1940

RESULT 13

AAR12599
ID AAR12599 standard; Protein; 1250 AA.

XX AAR12599;

XX 25-MAR-2003 (updated)

XX 17-SEP-1991 (first entry)

XX Portion of PT-NANBH viral non-structural protein.

XX post-transfusional non-A, non-B hepatitis; virus; vaccine; ss.

XX Non-A, non-B hepatitis virus.

XX GB2239245-A.

XX 26-JUN-1991.

XX 17-DEC-1990; 90GB-0027250.

XX 03-MAR-1990; 90GB-0004814.

XX 18-DEC-1989; 89GB-0028562.

XX 27-FEB-1990; 90GB-0004414.

XX 17-DEC-1990; 90GB-0027250.

XX (WELL) WELLCOME FOUND LTD.

XX (HIGH) HIGHFIELD P E.

XX Highfield PE, Rodgers BC, Tedder RS, Barbara JAJ;

XX WPI; 1991-187584/26.

XX N-PSDB; AAQ12241.

XX Post-transfusional non-A non-B hepatitis polypeptide(s) - and
PT also DNA and antibodies used in diagnostic assays and in vaccines
XX Claim 1; Page 88-97; 108pp; English.
XX The sequence was deduced from a non-structural (3') coding region
CC sequence isolated from serum of humans infected by the PT-NANBH

This sequence is encoded by the genome of a blood transmissible non-A, non-B hepatitis (NANBH) virus. The cDNA sequence was isolated using the primers given in AA063500-35. The amplified fragments are used in the detection of hepatitis virus. The target DNA was isolated from serum of chronically infected NANBH patients who were C100 antibody-positive and HCV RNA (NS5 region) positive. Reverse transcription-PCR and PCR were performed on cDNA and the total human NANBH DNA was constructed from 23 clones.

Sequence 3010 AA;

Query Match 14.1%; Score 257; DB 15; Length 3010;
Best Local Similarity 31.5%; Pred. No. 1.2e-12;
Matches 84; Conservative 31; Mismatches 74; Indels 78; Gaps 10;

QY 23 RPAVTPDREVLYQEFDEMEE-ASHLPYIEQGMQLAEQFKOKALAF---ASRQNHVSPPTHY 78
DB 1691 RPAVTPDREVLYQEFDEMEECASHLPYIEQGMQLAEQFKOKALGLQTATKQAE----- 1742
79 VPESDARPAIIPDREVLYHREFDEMEEAFASR-GNHVSPAHYVPESDASQAAPYIEQ----- 133
1743 ---AEAAPVVESE-----WRALEAFWAKHMMNFIISGIOYLAGLSTLPGNPAIASLMAF 1793
QY 134 -AQVIAHQFKEKVLAF-----ASRGNHDSPTHYVPESDAKPPI----- 170
DB 1794 TASITSPITONTLLFNILGWVAAQLAPPSAASXFGAGIAGALISIGLKVLDILA 1853
QY 171 -----IPDREVLYREFDEMEESQHL--PYIEQGMML-----AE 201
DB 1854 GYGAGVAGALVAFKVMGSEAPSAEDLVNLLPALISPGALVXXVCAILRRHVGPGEAV 1913
QY 202 QFKOKALAFASRGNHVPPTHYVETSDA 228
DB 1914 QMMNRLIAPASRGNHVSPTHYVPESDA 1940

RESULT 15

AA:68622
ID AAR68622 standard; Protein: 3010 AA.

AC AAR68622;

DT 16-OCT-1995 (first entry)

DE HCV protein cleavable with new serine proteinase.

KW proteinase; serine; cleavage; hepatitis C virus; HCV.

Hepatitis C Virus.

FH Key Location/Qualifiers
FT Cleavage-site 2419..2420
FT /note="Serine protease cleavage site"

PN JP06315377-A.

PD 15-NOV-1994.

PF 06-MAY-1993; 93JP-0105666.

PR 06-MAY-1993; 93JP-0105666.

PA (KAEN/) KAENNO K.

PA (SOYA-) SOYAKU GIUTSU KENKYUSHO KK.

PA (SUMO) SUMITOMO METAL IND LTD.

DR N-PSDB; AAQ80498.

PT New HCV-originated proteinase active substance - used for
PT site-specific cleavage by an intermolecular reaction and the
PT purification thereof

PS Disclosure; Page 10-19; 23pp; Japanese.

XX This protein from HCV (hepatitis C virus) (encoded by AA080498) is
CC cleaved between amino acids 2419 and 2420, by a new serine protease,
CC contg. the sequence of AAR68621. The proteinase is purified as a fused
CC product with the dihydrofolate reductase protein by using a methotrexate
CC column. It can be used for the development of an inhibitor for HCV
CC proteinase.

Sequence 3010 AA;

Query Match 14.1%; Score 257; DB 16; Length 3010;
Best Local Similarity 31.8%; Pred. No. 1.2e-12;
Matches 85; Conservative 29; Mismatches 75; Indels 78; Gaps 10;

QY 23 RPAVTPDREVLYQEFDEMEE-ASHLPYIEQGMQLAEQFKOKALAF---ASRQNHVSPPTHY 78
DB 1691 RPAVTPDREVLYQEFDEMEECASHLPYIEQGMQLAEQFKOKALGLQTATKQAE----- 1745
79 VPESDARPAIIPDREVLYHREFDEMEEAFASR-GNHVSPAHYVPESDASQAAPYIEQ----- 133
1746 ---AAP-----VESKWRALBFVWAKHMMNFIISGIOYLAGLSTLPGNPAIASLMAF 1793
QY 134 -AQVIAHQFKEKVLAF-----ASRGNHDSPTHYVPESDAKPPI----- 170
DB 1794 TASITSPITONTLLFNILGWVAAQLAPPSAASAFVGAGIAGAAVGSIGLKVLDILA 1853
QY 171 -----IPDREVLYREFDEMEESQHL--PYIEQGMML-----AE 201
DB 1854 GYGAGVAGALVAFKVMGSEAPSTEDLVNLLPALISPGALVGVCAAILRRHVGPGEAV 1913
QY 202 QFKOKALAFASRGNHVPPTHYVETSDA 228
DB 1914 QMMNRLIAPASRGNHVSPTHYVPESDA 1940

Search completed: July 25, 2003, 13:09:44
Job time : 87 secs